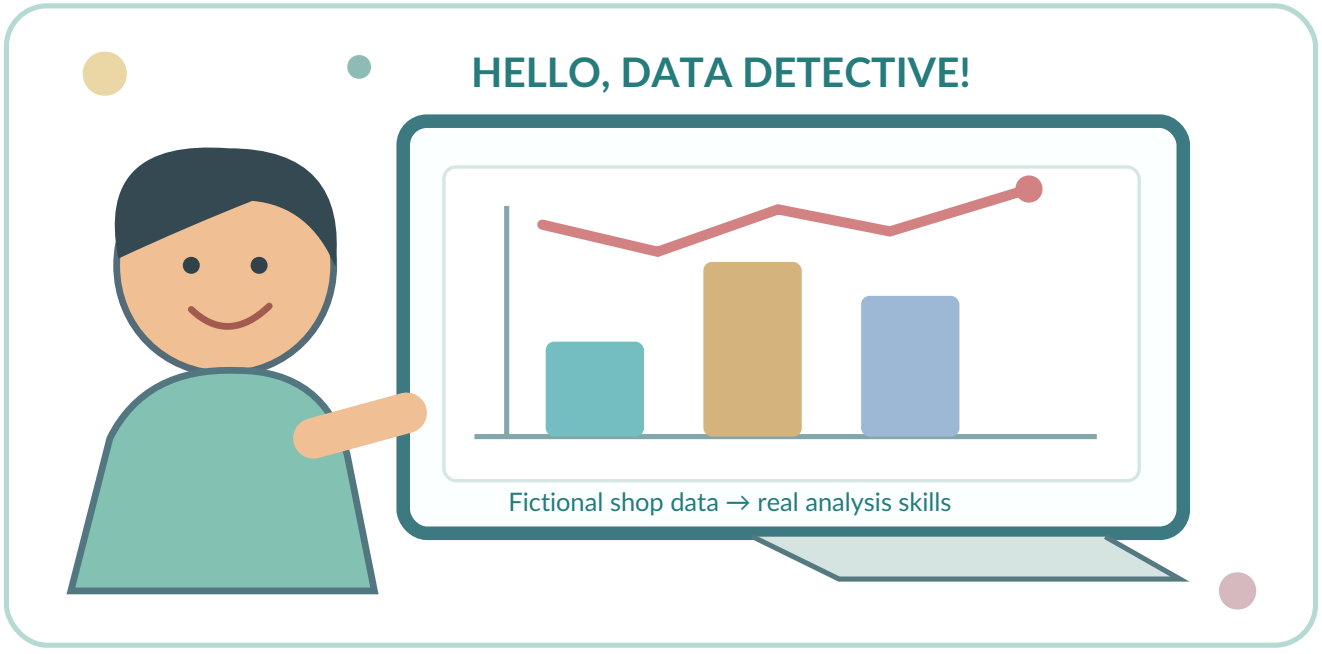
EGOTECHWORLD • KIDS DATA LAB • 2026

Data Analysis
මුල සිට ඉගෙනගමු!

Python • pandas • Charts • Reports

වයස අවුරුදු 10 සිට ඉගෙනගන්නා දරුවන්ටත් ගැළපෙන සිංහල පියවරෙන් පියවර පොතක්.
Little Star Shop හි දත්ත කතාව අලුතින් සොයාගමු!



WHITE PAGES | VS CODE | WINDOWS CMD | 2026

විස්තර • Source code • Real charts • කුඩා challenges • MCQ • Final report
egotechworld.com

HOW TO USE THIS BOOK

මේ පොත කියවන්න ඉගෙනගන්නේ කොහොමද?

මෙහි සෑම පාඩමක්ම එක් කුඩා වැඩක් ඉලක්ක කරනවා. code එක කියවන්න, CMD preview එක බලන්න, ඔබගේ laptop එකේ run කරලා අභ්‍යාසය කරන්න. එය නිවැරදිව වැඩ කරන විට ඊළඟ පාඩමට යන්න.

1. උදාහරණය කියවා අදහස තේරුම් ගන්න
2. VS Code panel එකේ code අනුගමනය කරන්න
3. CMD එකේ run කර output පරීක්ෂා කරන්න



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

◆ Visual Studio Code / lessons/00_hello.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
print("Hello, Data Detective!")
```

■ Windows Command Prompt / example output

```
C:\> python lessons/00_hello.py
Hello, Data Detective!
```

MINI MISSION

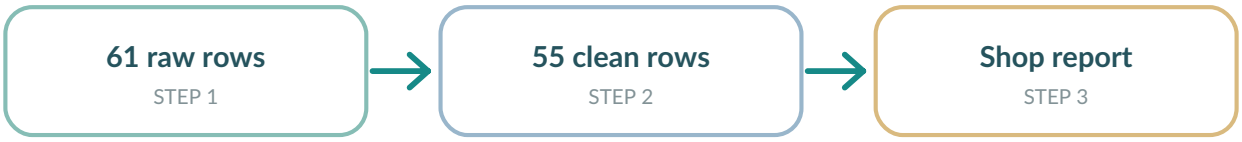
ඔබගේ computer එකේ ඉගෙනීමට ගත හැකි දින 7ක් තෝරා සටහන් කරන්න.

OUR PRACTICE PROJECT

අපේ Little Star Shop

අපේ කඩය සම්පූර්ණයෙන් fictional! පාරිභෝගික නමක් හෝ රහස් දත්ත භාවිතා නොකර, 2026 සැප්තැම්බර් sample sales table එකෙන් වටිනා දේවල් දැනගන්නවා.

Order ID	Date	Product	Qty	Unit (Rs)
S001	2026-09-01	Pencil	1	40
S002	2026-09-01	Notebook	2	180
S003	2026-09-01	Eraser	3	30
S004	2026-09-02	Water Bottle	4	350



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

අපි සොයාගන්න ප්‍රශ්න

වැඩිම revenue දුන් category එක කුමක්ද? මුළු units කීයද? Data වල වැරදි values කොපමණද? Chart වලින් තේරුම් ගන්න පුළුවන් සීමාව කුමක්ද?

> **FACT CHECK**
>
> මෙම sample dataset එකේ clean sale lines 55, sold units 135, revenue Rs 20,120.00. මෙය profit අගයක් නොවේ.

THE COMPLETE LEARNING PATH

අන්තර්ගතය • Units 01–09

Python basics සිට data cleaning, groupby, charts, reports සහ 2026 AI-era data thinking දක්වා.

UNIT 01 • Data Detective
දත්ත කියවීමේ රහස්
7 lessons + quiz

UNIT 02 • Python Basics
Python මුල සිට
11 lessons + quiz

UNIT 03 • Files and Setup
CSV සහ project setup
9 lessons + quiz

UNIT 04 • pandas Tables
DataFrame හඳුනාගනිමු
10 lessons + quiz

UNIT 05 • Data Cleaning
වැරදි දත්ත සකස් කරමු
12 lessons + quiz

UNIT 06 • Simple Statistics
ගණන් සහ groupby
10 lessons + quiz

UNIT 07 • Charts and Stories
Charts හදා කියවමු
9 lessons + quiz

UNIT 08 • Reports and Tests
Reports සහ checks
8 lessons + quiz

UNIT 09 • 2026 Data Literacy
AI කාලයේ Data Thinking
8 lessons + quiz

KEEP PRACTICING

එක් දවසකට පාඩම් 2–3ක් කිරීම හොඳයි. පිටු ගණනට වඩා, ක්‍රියාවලිය ඔබම කර බැලීම වැදගත්.

A LITTLE GUIDE FOR GROWN-UPS

දෙමාපියන් සහ ගුරුවරුන්ට

මේ පාඩම තාක්ෂණිකව නිවැරදිව තිබුණත් ඉතා සරල සංකල්පවලින් ආරම්භ වෙනවා. Install සහ file permissions වැනි වැඩේදී වැඩිහිටි සහය ලබාදීම සුදුසුයි.

- ☐ Python, VS Code සහ packages install කරන විට නිල වෙබ් අඩවි භාවිතා කරන්න.
- ☐ Project එකේ synthetic (මනඃකල්පිත) sales data පමණක් ඇත.
- ☐ සැබෑ පාරිභෝගික නම්, ලිපින, දුරකථන, ගෙවීම් දත්ත එක් නොකරන්න.
- ☐ “Revenue” සහ “Profit” වෙන්කර පැහැදිලි කරන්න.
- ☐ දරුවා code copy කරනවාට අමතරව “ඇයි?” කියා අසන බව දිරිගන්වන්න.
- ☐ 2026 tools වෙනස් විය හැකි නිසා official documentation යොදා තහවුරු කරන්න.

◆ Visual Studio Code / **CMD • main setup**

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
python run_all.py
python -m unittest discover -s tests -v
```

SAFETY

මෙය offline educational dataset එකකි. Live store, finance system, හෝ third-party customer database එකකට සම්බන්ධ නොවේ.

EGOTECHWORLD • DATA ANALYSIS KIDS LAB • 2026

01

දත්ත කියවීමේ රහස්

Data Detective

දත්ත යනු මොනවාද, නිවැරදි ප්‍රශ්න අසන්නේ කෙසේද?



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

WHAT YOU WILL LEARN

7 කුඩා පාඩම්, code examples සහ නිවැරදි data reasoning

TRY IT YOURSELF

සෑම පිටුවකම mini mission. Unit අවසානයේ MCQ ත්‍රිත්වයක්.

> දත්ත කියවන්න. සාක්ෂි පරීක්ෂා කරන්න. අවධානයෙන් නිගමනය කරන්න.

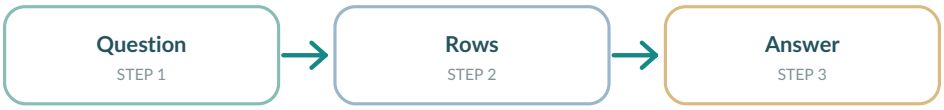
UNIT 01 / LESSON 01

Data Analysis කියන්නේ මොකක්ද?

Data Analysis කියන්නේ බොහෝ අංක සහ වගු අතරින් ප්‍රයෝජනවත් ප්‍රශ්නවලට සාක්ෂි සහිත පිළිතුරු සොයාගැනීමයි. අපේ මනඃකල්පිත Little Star Shop එකේ වැඩිපුර විකුණෙන භාණ්ඩය කුමක්ද කියා අද සොයමු.

WATCH THE STEPS

1. Question
2. Collect data
3. Find an answer



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / lessons/00_hello.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
print("Hello, Data Detective!")
print("Let the numbers tell a story")
```

■ Windows Command Prompt / example output

```
C:\> python lessons/00_hello.py
Hello, Data Detective!
We will discover stories hidden inside numbers.
```

REMEMBER

Data Analysis = සාක්ෂි මත තීරණ; අනුමාන හෝ පුද්ගලික මතයක් පමණක් නොවේ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

ඔබේ පන්තිය ගැන දත්ත භාවිතයෙන් පිළිතුරු දිය හැකි ප්‍රශ්න දෙකක් ලියන්න.

UNIT 01 / LESSON 02

අපේ Little Star Shop project එක

මේ පොතේ සැබෑ ගනුදෙනු භාවිතා කරන්නේ නැහැ. 2026 සැප්තැම්බර් මාසයේ සිදු වූ බවට අපි නිර්මාණය කළ sales records විශ්ලේෂණය කරනවා. Pencil, Notebook, Eraser, Water Bottle, Juice, Biscuit භාණ්ඩ වේ.

WATCH THE STEPS

1. Meet our shop
2. See sales table
3. Ask questions



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / lessons/02_lists.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
products = ["Pencil", "Notebook", "Eraser"]
print("First product:", products[0])
```

■ Windows Command Prompt / example output

```
C:\> python run_all.py
Clean rows    : 55
Revenue (not profit): Rs 20,120.00
```

REMEMBER

මේවා උදාහරණ ගණන් පමණයි. ඒවා වෙළඳපොළේ සැබෑ මිල නොවේ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

ඔබේ කඩයට අලුත් භාණ්ඩ දෙකක් යෝජනා කර ඒවාට මිල නියම කරන්න.

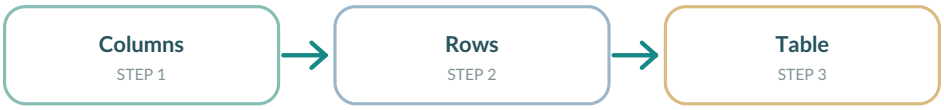
UNIT 01 / LESSON 03

Data table එක කියවමු

Table එකක row එකක් කියන්නේ එක් වාර්තාවක්. Column එකක් කියන්නේ එම වාර්තාවේ එකම වර්ගයේ තොරතුරක්. අපේ row එකකට order_id, date, product, category, quantity, unit_price තිබෙනවා.

WATCH THE STEPS

1. Look at headings
2. Read a row
3. Read a column



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

CSV / TABLE EXAMPLE

◆ Visual Studio Code / notes/table_example.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
row = {"product": "Pencil", "quantity": 2, "unit_price": 40}
print(row["product"])
```

■ Windows Command Prompt / example output

```
order_id  date        product  qty  price
S001      2026-09-01  Pencil   1    40
S002      2026-09-01  Notebook 2    180
```

REMEMBER

Row එකක් = එක් sale line එකක්. එකම customer කෙනෙකුගේ සියලු මිලදී ගැනීම් ලෙස මෙය නොසලකන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Order ID සහ Product යන columns දෙක එකිනෙකට වෙනස් අර්ථ දෙන්නේ කොහොමද?

UNIT 01 / LESSON 04

Data → Information → Decision

කඩයක පරිගණකයට දිනපතා sales rows ලැබේ. ඒවා එකතු කර “Accessories වැඩි ආදායම දෙන category එක” කියා දැනගත හැක. නමුත් stock වැඩිකිරීම වැනි තීරණයකට මිල, වියදම් සහ දත්ත කාල සීමාවත් බලන්න ඕන.

WATCH THE STEPS

1. Collect
2. Analyze
3. Decide carefully



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / notes/decision.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
revenue_by_type = {"Accessories": 10500, "Stationery": 5000}
print(max(revenue_by_type, key=revenue_by_type.get))
```

■ Windows Command Prompt / example output

```
Accessories : Rs 10,500
Stationery  : Rs  5,000
Snacks      : Rs  4,620
```

REMEMBER

Income වැඩි category එක profit වැඩි category එක විය යුතුම නැහැ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

එක මාසයක ආදායම පමණක් දැනගෙන වසරකට අදාළ තීරණයක් ගත හැකිද? ඇයි?

UNIT 01 / LESSON 05

ප්‍රශ්නයක් හරි හැටි අසමු

“අපේ කඩය හොඳද?” කියන ප්‍රශ්නය විශාලයි. “සැප්තැම්බර් sales records අතර වැඩිම ආදායම ගෙනා product එක කුමක්ද?” කියන ප්‍රශ්නයට table එකෙන් පිළිතුරු සොයාගන්න පුළුවන්.

WATCH THE STEPS

1. Ask
2. Measure
3. Check



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / notes/questions.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
question = "Which product has the most revenue?"
print(question)
```

■ Windows Command Prompt / example output

```
Good question: Which category earned
the most revenue in these sample rows?
```

REMEMBER

සෑම graph එකක්ම පළමුව පැහැදිලි ප්‍රශ්නයකට සම්බන්ධ විය යුතුයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

අපැහැදිලි ප්‍රශ්නයක් තෝරා ඒක data වලින් පිළිතුරු දිය හැකි ලෙස නැවත ලියන්න.

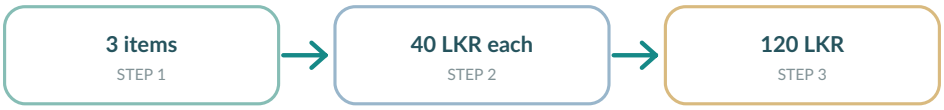
UNIT 01 / LESSON 06

Revenue සහ Profit වෙනස

Revenue කියන්නේ විකුණුම්වලින් ලැබුණු මුළු මුදල. Profit ගැන කතාකරන්න භාණ්ඩ මිලදී ගැනීමේ වියදම්, ගාස්තු සහ වෙනත් වියදම්ත් දැනගන්න ඕන. අපේ dataset එකේ ඒ වියදම් නැති නිසා profit ගණනය කළ නොහැක.

WATCH THE STEPS

1. Quantity
2. Price
3. Revenue



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / lessons/01_variables.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
quantity = 3
unit_price = 40
revenue = quantity * unit_price
print("Rs", revenue)
```

■ Windows Command Prompt / example output

```
C:\> python lessons/01_variables.py
Pencil : Rs 120
```

REMEMBER

Revenue = quantity × unit price. එය Profit නොවේ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Pencil 4ක් රු.40 බැගින් විකුණූ විට Revenue කීයද?

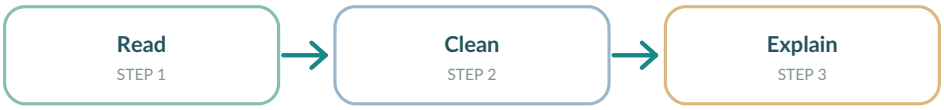
UNIT 01 / LESSON 07

2026 දත්ත ඉගෙනීමේ මාර්ගය

දත්ත විශ්ලේෂණයට Excel/Sheets, SQL, Python වැනි ක්‍රම විවිධව භාවිතා කරනවා. අපි Python + pandas + Matplotlib මාර්ගය තෝරනවා. AI tool එකක් code යෝජනා කළත් අගයන් සෑමවිටම පරීක්ෂා කරන්න.

WATCH THE STEPS

1. Python
2. pandas
3. Charts



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / notes/tools.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
TOOLS = ["Python", "pandas", "Matplotlib"]
for tool in TOOLS:
    print(tool)
```

■ Windows Command Prompt / example output

```
Python
pandas
Matplotlib
```

> **REMEMBER**
> මේ පොතේ අපි data විස්තර කරනවා; Machine Learning model train කරන්නේ නැහැ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Data Analysis සහ Machine Learning එකම දේ නොවන බව වාක්‍යයකින් පැහැදිලි කරන්න.

UNIT 01 • THINK AND CHOOSE

කුඩා දැනුම් පරීක්ෂාව

වැරදුණොත් පසුගිය පාඩම නැවත කියවන්න. Answers අවසානයේ තිබෙනවා; මුලින්ම ඔබම උත්සාහ කරන්න.

1. Data analysis එකේ ප්‍රධාන අරමුණ?

A. අංක අලංකාර කිරීම
B. සාක්ෂි සහිත පිළිතුරු සෙවීම
C. Guess කිරීම

2. Revenue එක profit ද?

A. ඔව්
B. නැහැ
C. සෑමවිටම

3. Table එකක row යනු?

A. එක් වාර්තාවක්
B. හැම column එකක්
C. Chart එකක්

කඩදාසියේ පිළිතුරු දෙන්න

1. මෙම Unit එකේ ඉගෙනගත් වැදගත්ම අදහස කුමක්ද?
2. මේ සතියේ ඔබම run කළ Python code එක කුමක්ද?
3. ඊළඟට වැඩිදුර පැහැදිලි කරගන්න අවශ්‍ය ප්‍රශ්නය කුමක්ද?

MINI MISSION

මිතුරෙකුට මිනිත්තු දෙකකින් අද පාඩමේ අදහස පැහැදිලි කරන්න.

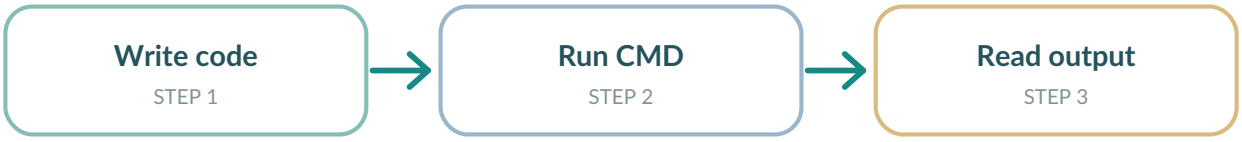
EGOTECHWORLD • DATA ANALYSIS KIDS LAB • 2026

02

Python මුල සිට

Python Basics

Variables, loops, functions සහ අංක ගණනය කිරීම.



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

WHAT YOU WILL LEARN

11 කුඩා පාඩම්, code examples සහ නිවැරදි data reasoning

TRY IT YOURSELF

සෑම පිටුවකම mini mission. Unit අවසානයේ MCQ ත්‍රිත්වයක්.

> දත්ත කියවන්න. සාක්ෂි පරීක්ෂා කරන්න. අවධානයෙන් නිගමනය කරන්න.

UNIT 02 / LESSON 01

Python සහ VS Code සූදානම් කරමු

Python 3.11 හෝ අලුත් supported version එකක් සහ VS Code install කරමු. Install කරන විට විශ්වාසදායක නිල websites භාවිතා කරන්න. Windows CMD එකේ Python version පෙන්වෙන්නේද කියා බලමු.

WATCH THE STEPS

1	2	3
Install	Open CMD	Check version



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / CMD • setup

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
py --version
python --version
```

■ Windows Command Prompt / example output

```
C:\> py --version
Python 3.x.x
```

REMEMBER

Windows දී `py` නැතිනම් ක්‍රියාකරන `python` command එක භාවිතා කරන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

ඔබේ computer එකේ Python version එක ලියාගන්න.

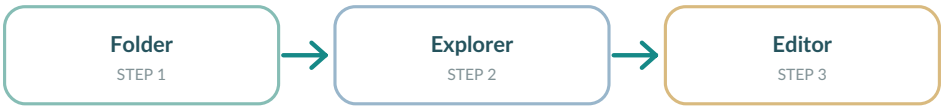
UNIT 02 / LESSON 02

Project folder එක හදමු

File එක හොයාගන්න පහසු වීමට හැමදේම `data_analysis_kids_sinhala_2026` folder එකේ තබමු. VS Code → File → Open Folder මඟින් folder එක open කර Terminal → New Terminal තෝරන්න.

WATCH THE STEPS

1. Unzip
2. Open folder
3. See files



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / CMD • folder

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
cd data_analysis_kids_sinhala_2026
code .
```

■ Windows Command Prompt / example output

```
C:\> cd data_analysis_kids_sinhala_2026
C:\data_analysis_kids_sinhala_2026>
```

REMEMBER

`code .` command එක නොදන්නා නම් VS Code මෙනුවෙන් Open Folder භාවිතා කරන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

`data/`, `lessons/`, `outputs/` folders තුනේ වැඩ මොකක්ද කියා ලියන්න.

UNIT 02 / LESSON 03

print() - තිරයට ලියමු

print() කියන Python function එකෙන් අපට වචන සහ අංක තිරයට පෙන්වන්න පුළුවන්. Quotes ඇතුළේ වචන ලියන විට ඒවා text ලෙස සලකනවා.

WATCH THE STEPS

1. Write
2. Save
3. Run



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / lessons/00_hello.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
print("Hello, Data Detective!")
print("My first analysis is starting")
```

■ Windows Command Prompt / example output

```
C:\> python lessons/00_hello.py
Hello, Data Detective!
```

REMEMBER

Code save කිරීමෙන් පසු නැවත run කළ යුතුයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

"I love charts" කියලා තවත් line එකක් print කරන්න.

UNIT 02 / LESSON 04

Variables - නම් දාපු පෙට්ටි

Variable කියන්නේ අගයක් තබාගන්න නමක්. product කියන නමෙන් text එකක්ද, quantity කියන නමෙන් number එකක්ද තබන්න පුළුවන්. = මඟින් අගය assign කරනවා.

WATCH THE STEPS

1	2	3
Name values	Calculate	Print



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / lessons/01_variables.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
product = "Pencil"
quantity = 3
unit_price = 40
print(product, quantity * unit_price)
```

■ Windows Command Prompt / example output

```
C:\> python lessons/01_variables.py
Pencil : Rs 120
```

REMEMBER

Python හි = assignment; == comparison.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

quantity = 5 කර Revenue කීය වෙයිද කියා ධාවනය කිරීමට පෙර ගණනය කරන්න.

UNIT 02 / LESSON 05

Data types: str, int, float

"120" කියන්නේ text (str); 120 කියන්නේ integer (int); 120.50 කියන්නේ decimal (float). ගණනයේදී number එකක් අවශ්‍ය නම් text එකෙන් නිකම්ම ගුණ කරන්න බැහැ.

WATCH THE STEPS

1. Strings
2. Integers
3. Floats



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / notes/types.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
name = "Notebook"
count = 2
price = 180.0
print(type(name), type(count), type(price))
```

■ Windows Command Prompt / example output

```
<class str>  <class int>  <class float>
```

REMEMBER

CSV එකකින් ආ data බොහෝවිට text ලෙස ආරම්භ වේ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

“50” සහ 50 අතර වෙනස කුමක්ද?

UNIT 02 / LESSON 06

Lists - අගයන් කීපයක්

List එකේ අගයන් [] ඇතුළේ comma වලින් වෙන් කළ හැක. len() භාවිතයෙන් ගණන ගනිමු. sum() මඟින් අගයන් එකතු කරමු.

WATCH THE STEPS

1. Collect numbers
2. Count
3. Sum



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / lessons/02_lists.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
sales = [120, 250, 80, 350]
print(len(sales))
print(sum(sales))
```

■ Windows Command Prompt / example output

```
C:\> python lessons/02_lists.py
Count: 4
Total: 800
Average: 200.0
```

REMEMBER

List එකේ index 0 පළමුව item එකයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

තවත් sale එකක් ලෙස 100 එකතු කර total නැවත හදාගන්න.

UNIT 02 / LESSON 07

Average - මැද අගයක් ගැන

Mean average එක ගන්නේ එකතුව අගයන් ගණනෙන් බෙදලා. 120,250,80,350 නම් එකතුව 800; count 4; mean 200. විශාල අගයක් mean එක ඉහළට ඇද ගන්න පුළුවන්.

WATCH THE STEPS

1. Add
2. Divide
3. Interpret



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

Visual Studio Code / lessons/02_lists.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
sales = [120, 250, 80, 350]
average = sum(sales) / len(sales)
print(average)
```

Windows Command Prompt / example output

```
Average: 200.0
```

REMEMBER

Mean එක හැම data value එකක්ම සම්පූර්ණයෙන් නියෝජනය නොකරයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

1000ක් අලුතින් list එකට එක් කළොත් mean එක වෙනස් වන්නේ කොහොමද?

UNIT 02 / LESSON 08

for loop - හැම එකක්ම බලමු

List එකේ සෑම price එකක්ම print කරන්න පේළි තුනක් නැතුව for loop එක භාවිතා කරමු. Colon එක සහ indentation Python හි වැදගත්.

WATCH THE STEPS

1	2	3
Pick item	Print	Repeat



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / lessons/03_loops.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
prices = [40, 180, 30]
for price in prices:
    print("Price: Rs", price)
```

■ Windows Command Prompt / example output

```
Price: Rs 40
Price: Rs 180
Price: Rs 30
```

REMEMBER

Python හි `for` loop තුළ spaces 4ක් සාමාන්‍ය රටාවකි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Loop එකට තවත් price එකක් එක්කර total manually සොයන්න.

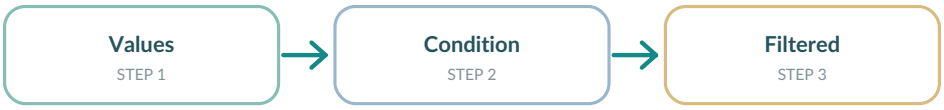
UNIT 02 / LESSON 09

if statement - පෙරහනක්

කැමති මිලකට වඩා වැඩි items පමණක් පෙන්වීමට if එකක් යොදන්න පුළුවන්. මෙය pandas filtering සඳහා හොඳ මුලික අදහසක්.

WATCH THE STEPS

1. Check price
2. True?
3. Display



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

Visual Studio Code / notes/filter_prices.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
prices = [40, 180, 30, 350]
for price in prices:
    if price > 100:
        print(price)
```

Windows Command Prompt / example output

```
180
350
```

REMEMBER

> වැඩිය; >= වැඩිය හෝ සමාන.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

`> 100` වෙනුවට `>= 100` දාලා වෙනස පැහැදිලි කරන්න.

UNIT 02 / LESSON 10

Functions - නැවත භාවිතා කරන්න

එකම calculation එක නැවත නැවත ලියන්න ඕන නැතිව function එකක් හදමු. Input arguments දෙකක් අරගෙන Revenue return කරනවා.

WATCH THE STEPS

1 Define

2 Call

3 Return



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / lessons/04_functions.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
def calculate_revenue(quantity, unit_price):
    return quantity * unit_price

print(calculate_revenue(2, 180))
```

■ Windows Command Prompt / example output

```
C:\> python lessons/04_functions.py
360
```

REMEMBER

මෙය revenue පමණයි, profit ගණනයක් නොවේ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

`calculate_revenue(4, 90)` හි පිළිතුර කීයද?

UNIT 02 / LESSON 11

Error හඳුනාගැනීම

Code එක run නොවුණාම කලබල නොවන්න. Error message එකේ file name, line number සහ වැරැද්ද කියවමු. Comma, colon, quotes හා spaces වගේ දේවල් පළමුව පරීක්ෂා කරන්න.

WATCH THE STEPS

1. Read error
2. Find line
3. Fix



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / notes/error_example.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
value = "120"
print(int(value) + 30)
# int() turns numeric text into a number.
```

■ Windows Command Prompt / example output

```
C:\> python notes/error_example.py
150
```

REMEMBER

Error එකක් යනු ඉගෙනීමට clue එකක්; නොදන්නා code copy කරන්න එපා.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

`int("hello")` දුන්නොත් මොකද වෙන්නේ? හේතුව කියන්න.

UNIT 02 • THINK AND CHOOSE

කුඩා දැනුම් පරීක්ෂාව

වැරදුණොත් පසුගිය පාඩම නැවත කියවන්න. Answers අවසානයේ තිබෙනවා; මුලින්ම ඔබම උත්සාහ කරන්න.

1. Python හි screen එකට ලියන්නේ?

A. plot()
B. print()
C. sum()

2. List එකක් සාදන්නේ?

A. []
B. {}
C. <>

3. 4 සහ 90 ගුණිතය?

A. 94
B. 360
C. 45

කඩදාසියේ පිළිතුරු දෙන්න

1. මෙම Unit එකේ ඉගෙනගත් වැදගත්ම අදහස කුමක්ද?
2. මේ සතියේ ඔබම run කළ Python code එක කුමක්ද?
3. ඊළඟට වැඩිදුර පැහැදිලි කරගන්න අවශ්‍ය ප්‍රශ්නය කුමක්ද?

MINI MISSION

මිතුරෙකුට මිනිත්තු දෙකකින් අද පාඩමේ අදහස පැහැදිලි කරන්න.

EGOTECHWORLD • DATA ANALYSIS KIDS LAB • 2026

03

CSV සහ project setup

Files and Setup

XAMPP අවශ්‍ය නැහැ. Python, VS Code, CSV සහ libraries.



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

WHAT YOU WILL LEARN

9 කුඩා පාඩම්, code examples සහ නිවැරදි data reasoning

TRY IT YOURSELF

සෑම පිටුවකම mini mission. Unit අවසානයේ MCQ ත්‍රිත්වයක්.

> දත්ත කියවන්න. සාක්ෂි පරීක්ෂා කරන්න. අවධානයෙන් නිගමනය කරන්න.

UNIT 03 / LESSON 01

CSV කියන්නේ මොකක්ද?

CSV = Comma-Separated Values. Spreadsheet එකක් වගේ columns සහ rows text file එකක ගබඩා කරයි. පළමු line එකේ headings; ඉතිරි line වල sales records තබනවා.

WATCH THE STEPS

1. Headers
2. Rows
3. Commas



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

CSV / TABLE EXAMPLE

◆ Visual Studio Code / data/shop_sales_raw.csv

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
order_id,date,product,quantity
S001,2026-09-01,Pencil,1
S002,2026-09-01,Notebook,2
```

■ Windows Command Prompt / example output

```
order_id  date        product  qty
S001      2026-09-01  Pencil   1
S002      2026-09-01  Notebook 2
```

REMEMBER

CSV එක text file එකක්; extension `.csv` පමණක් වෙනස්.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

දෙවන row එකේ product නම සහ quantity හඳුනාගන්න.

UNIT 03 / LESSON 02

CSV sample එක සාදමු

ZIP එකේ `data/create_sample_data.py` file එකෙන් fictional sales data සාදනවා. එම script එක run කළ විට `data/shop_sales_raw.csv` ලැබෙනවා. අලුත් customer තොරතුරු ඇතුළත් නොකරමු.

WATCH THE STEPS

1. Run generator
2. Open CSV
3. Read headings



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / CMD • generator

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
python data/create_sample_data.py
```

■ Windows Command Prompt / example output

```
C:\> python data/create_sample_data.py
Created: ...shop_sales_raw.csv
```

REMEMBER

Dataset එක සම්පූර්ණයෙන් fictional; දරුවන්ගේ පෞද්ගලික දත්ත නොමැත.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

CSV එකේ headings 6 ලියාගන්න.

UNIT 03 / LESSON 03

CSV එක text editor එකේ විවෘත කරන්න

VS Code වල CSV file එක open කළොත් comma වලින් වෙන් වූ අගයන් පෙනෙයි. අගයන් අතර හිස් values, අවකාශ, සහ typo තිබෙන්න පුළුවන්. ඒවා හිතුවක්කාරව මාරු නොකරමු.

WATCH THE STEPS

1. Open CSV
2. Check row
3. Keep raw



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

CSV / TABLE EXAMPLE

◆ Visual Studio Code / data/shop_sales_raw.csv

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
order_id,date,product,category,quantity,unit_price
S001,2026-09-01,Pencil,Stationery,1,40
```

■ Windows Command Prompt / example output

```
S001,2026-09-01,Pencil,Stationery,1,40
```

REMEMBER

Raw data file එක වෙනස් නොකර cleaned output එක වෙනම save කිරීම හොඳ ක්‍රමයකි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

`S001` row එකේ ආදායම කීයද?

UNIT 03 / LESSON 04

Virtual environment (.venv)

එක් project එකක libraries වෙනත් projects වලින් වෙන් කරන්න `.venv` හදමු. එය package install කිරීමේ ගැටලු අවම කරයි. වින්ඩෝස් CMD සඳහා පහත commands භාවිතා කරමු.

WATCH THE STEPS

1 Create

2 Activate

3 Install



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / CMD • venv

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
py -m venv .venv
.venv\Scripts\activate
```

■ Windows Command Prompt / example output

```
C:\> py -m venv .venv
C:\> .venv\Scripts\activate
(.venv) C:\>
```

REMEMBER

PowerShell/macOS/Linux activate commands README හි වෙනම ඇත.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

`.venv` folder එක GitHub වෙත upload නොකරන්නේ ඇයි?

UNIT 03 / LESSON 05

pandas සහ Matplotlib install

pandas මඟින් වගු සහ සංඛ්‍යා විශ්ලේෂණය කරනවා. Matplotlib charts නිර්මාණය කරනවා. `requirements.txt` කියන්නේ අවශ්‍ය packages වල ලැයිස්තුව.

WATCH THE STEPS

1. Read requirements
2. Install
3. Check



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / requirements.txt

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
python -m pip install -r requirements.txt
python -m pip show pandas
```

■ Windows Command Prompt / example output

```
(.venv) C:\> python -m pip show pandas
Name: pandas
```

REMEMBER

නිල pandas documentation අනුව pandas 2026 දී DataFrames, CSV, grouping ආදිය සඳහා භාවිතා කරයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

pandas නිවැරදිව install වූ බව CMD එකෙන් පරීක්ෂා කරන්න.

UNIT 03 / LESSON 06

path සහ file name ගැන

File path කියන්නේ file එකට යන මාර්ගය. `data/shop_sales_raw.csv` කියන්නේ project root එකේ ඇති data folder එක තුළ raw CSV එකයි. Filename එකේ ලොකු/කුඩා අකුරු වැදගත් විය හැක.

WATCH THE STEPS

1. Root folder
2. data folder
3. CSV



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

Visual Studio Code / 01_first_pandas.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
from pathlib import Path
ROOT = Path(__file__).resolve().parent
FILE = ROOT / "data" / "shop_sales_raw.csv"
print(FILE.exists())
```

Windows Command Prompt / example output

```
True
```

REMEMBER

File path වැරදි නම් FileNotFoundError පෙන්විය හැක.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Folder එක වෙනත් තැනකට move කළත් `ROOT` මඟින් file හොයාගැනීමේ වාසිය කුමක්ද?

UNIT 03 / LESSON 07

Python csv module එක

pandas ඉගෙනීමට පෙර Python හි `csv.DictReader` මඟින් CSV rows dictionary ලෙස කියවමු. මෙහි කිසිම extra library එකක් install කිරීමට අවශ්‍ය නැහැ.

WATCH THE STEPS

1. Open file
2. Read dictionary
3. Print



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / lessons/05_csv_reader.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
import csv
with open("data/shop_sales_raw.csv", encoding="utf-8") as f:
    reader = csv.DictReader(f)
    print(next(reader)["product"])
```

■ Windows Command Prompt / example output

```
C:\> python lessons/05_csv_reader.py
2026-09-01 Pencil 1
2026-09-01 Notebook 2
```

REMEMBER

Source script එක `Path` භාවිතයෙන් CSV location එක ස්ථිරව සොයාගනී.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

CSV row එකක් dictionary එකක් වගේ හිතන්න පුළුවන් ඇයි?

UNIT 03 / LESSON 08

Raw CSV එකේ වැරදි අගයන්

පිරිසිදු දත්ත පමණක් නැහැ. Sample එකේ හිස් quantity, `two` කියන quantity, අසත්‍ය date, 0 price, negative quantity සහ duplicate order එකක් තිබෙනවා. Data cleaning ඉගෙනගන්න ඒවා හිතාමතාම ඇතුළත් කරලා.

WATCH THE STEPS

1	2	3
Detect	Mark	Review



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / data/shop_sales_raw.csv

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
bad_quantity = "two"
print("Can this multiply?", bad_quantity)
```

■ Windows Command Prompt / example output

```
Raw rows      : 61
Rejected rows : 6
Clean rows    : 55
```

REMEMBER

වැරදි දත්ත වෙනස් කිරීමට පෙර “ඇයි මෙය වැරදිද?” යන්න තහවුරු කරන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

ඉහත වැරදි වර්ග හයෙන් දෙකක් තෝරා ඒවා ව්‍යාපාර තීරණයට බලපාන අයුරු කියන්න.

UNIT 03 / LESSON 09

Project එක මුල සිට run කරන්න

ZIP එක extract කර project root folder එකේ terminal එක open කර `python run_all.py` ධාවනය කරන්න. එය CSV සාදා, පිරිසිදු කර, වගු, charts, සහ text report එකක් සකස් කරනවා.

WATCH THE STEPS

1 Generate

2 Clean

3 Analyze



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / CMD • full project

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
python run_all.py
```

■ Windows Command Prompt / example output

```
C:\> python run_all.py
Clean rows    : 55
Revenue: Rs 20,120.00
Done! Open charts/ and outputs/.
```

REMEMBER

run_all.py එකම commands පිළිවෙළට ධාවනය කරයි; රහස් API key අවශ්‍ය නැහැ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

outputs folder එකේ නව files හතරක් හඳුනාගන්න.

UNIT 03 • THINK AND CHOOSE

කුඩා දැනුම් පරීක්ෂාව

වැරදුණොත් පසුගිය පාඩම නැවත කියවන්න. Answers අවසානයේ තිබෙනවා; මුලින්ම ඔබම උත්සාහ කරන්න.

1. CSV කියන්නේ?

A. Comma separated values
B. Cloud secure video
C. Code style view

2. venv එකේ අරමුණ?

A. Pictures
B. Project packages වෙන් කිරීම
C. Delete files

3. raw file ගැන හොඳ නීතිය?

A. Overwrite
B. Keep original
C. Share private data

කඩදාසියේ පිළිතුරු දෙන්න

1. මෙම Unit එකේ ඉගෙනගත් වැදගත්ම අදහස කුමක්ද?
2. මේ සතියේ ඔබම run කළ Python code එක කුමක්ද?
3. ඊළඟට වැඩිදුර පැහැදිලි කරගන්න අවශ්‍ය ප්‍රශ්නය කුමක්ද?

MINI MISSION

මිතුරෙකුට මිනිත්තු දෙකකින් අද පාඩමේ අදහස පැහැදිලි කරන්න.

EGOTECHWORLD • DATA ANALYSIS KIDS LAB • 2026

04

DataFrame හඳුනාගනිමු

pandas Tables

pandas මඟින් rows, columns සහ summary පරීක්ෂා කිරීම.



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

WHAT YOU WILL LEARN

10 කුඩා පාඩම්, code examples සහ නිවැරදි data reasoning

TRY IT YOURSELF

සෑම පිටුවකම mini mission. Unit අවසානයේ MCQ ත්‍රිත්වයක්.

> දත්ත කියවන්න. සාක්ෂි පරීක්ෂා කරන්න. අවධානයෙන් නිගමනය කරන්න.

UNIT 04 / LESSON 01

pandas සහ DataFrame

DataFrame කියන්නේ Python තුළ තිබෙන rows සහ columns ඇති table එකක්. pandas නිල tutorial වල data කියවීම, filter, group, statistics යනාදිය එහි ප්‍රධාන පාඩම්.

WATCH THE STEPS

1	2	3
Import	Create table	Inspect



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 01_first_pandas.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
import pandas as pd
print("pandas is ready")
```

■ Windows Command Prompt / example output

```
C:\> python 01_first_pandas.py
First five rows: ...
```

REMEMBER

`pd` කියන්නේ pandas සඳහා කෙටි නමක් පමණයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

DataFrame එකක් සහ paper table එකක් අතර සමානකම් දෙකක් ලියන්න.

UNIT 04 / LESSON 02

read_csv() - file එක කියවමු

CSV file එක DataFrame බවට පත් කිරීම `pd.read_csv()` මඟින් කරන්න පුළුවන්. File path නිවැරදිද කියා බලන්න. Raw data ගන්නා නිසා වැරදි values තවම තිබෙනවා.

WATCH THE STEPS

1. Read CSV
2. Create DataFrame
3. Display



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 01_first_pandas.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
import pandas as pd
sales = pd.read_csv("data/shop_sales_raw.csv")
print(sales.head())
```

■ Windows Command Prompt / example output

```
order_id       date    product     ...
S001      2026-09-01 Pencil      ...
```

REMEMBER

`read_csv()` එක data clean කරන්නේ නැහැ; මුලින්ම table එක කියවයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

ඔබේ code එකේ `sales` යන variable එකට වෙනත් සුදුසු නමක් යෝජනා කරන්න.

UNIT 04 / LESSON 03

head() සහ tail()

ඉතා දිගු table එකක් එකවර print කළොත් හොයාගන්න අමාරුයි. `head()` මුල් rows පහක් පෙන්වයි. `tail()` අවසාන rows පහක් පෙන්වයි.

WATCH THE STEPS

1	2	3
First rows	Last rows	Check



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 01_first_pandas.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
print(sales.head(3))
print(sales.tail(3))
```

■ Windows Command Prompt / example output

```
head(3): first 3 rows
tail(3): last 3 rows
```

REMEMBER

head() බලලා තනිවම සමස්ත දත්ත ගැන නිගමනය නොකරන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Raw CSV අවසානයේ duplicated row එකක් ඇත්තේ ඇයි කියා හඳුනාගන්න.

UNIT 04 / LESSON 04

shape - rows සහ columns

`sales.shape` කියන්නේ (rows, columns) දෙකක් දෙන tuple එකක්. අපේ raw CSV එකේ rows 61ක් සහ columns 6ක් තිබෙනවා. මේක dataset එකේ ප්‍රමාණය තේරුම්ගන්න උදව් වෙයි.

WATCH THE STEPS

1. Count rows
2. Count columns
3. Interpret



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 01_first_pandas.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
print(sales.shape)
print("Rows:", sales.shape[0])
print("Columns:", sales.shape[1])
```

■ Windows Command Prompt / example output

```
(61, 6)
Rows: 61
Columns: 6
```

REMEMBER

Rows ගණන විතරක් data quality මැනීමට ප්‍රමාණවත් නැහැ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Raw data 61 rows වුවත් clean data 55ක් වන්නේ මොන හේතුවකින්ද?

UNIT 04 / LESSON 05

columns සහ dtypes

`sales.columns` මඟින් headings බලන්න පුළුවන්. `sales.dtypes` මඟින් එක් column එකේ data type එක සොයාගත හැක. Quantity column එකේ `two` ඇති නිසා එය numeric නොවිය හැක.

WATCH THE STEPS

1. View names
2. Check types
3. Plan cleaning



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 01_first_pandas.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
print(sales.columns.tolist())
print(sales.dtypes)
```

■ Windows Command Prompt / example output

```
quantity    object
unit_price   int64
...
```

REMEMBER

Mixed text සහ numbers ඇති column එකක් සෘජුව sum කිරීමට පෙර clean කරන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Quantity වල `two` තිබීම නිසා mean() සෘජුව ගන්න ගැටලුවක් වන්නේ ඇයි?

UNIT 04 / LESSON 06

Select one column

`sales["product"]` මඟින් එක column එකක් තෝරගත හැක. එය Series යන pandas object එකකි. Value එකක් ගැන වැරදි තේරුමක් නොගන්න column heading බලන්න.

WATCH THE STEPS

1. Pick column
2. See values
3. Compare



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 01_first_pandas.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
print(sales["product"].head())
print(sales["category"].head())
```

■ Windows Command Prompt / example output

```
Pencil
Notebook
Eraser
```

REMEMBER

Square brackets තුළ column name එක නිවැරදිව තිබිය යුතුයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Product සහ Category දෙකේ වෙනස වචන කිහිපයකින් දක්වන්න.

UNIT 04 / LESSON 07

Select two columns

Columns කිහිපයක් තෝරන විට double square brackets භාවිතා කරනවා. මෙය කුඩා report table එකක් නිර්මාණය කිරීමට උපකාරීයි.

WATCH THE STEPS

1. Choose fields
2. Make view
3. Print



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 01_first_pandas.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
small = sales[["product", "quantity"]]
print(small.head(4))
```

■ Windows Command Prompt / example output

```
product       quantity
Pencil        1
Notebook      2
Eraser        3
```

REMEMBER

`[["A","B"]]` = columns දෙකක්, `["A"]` = එක column එකක්.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Date, Product, Quantity columns තුනම තෝරාගන්න code ලියන්න.

UNIT 04 / LESSON 08

iloc - row තෝරාගැනීම

`iloc[0]` මඟින් පළමු row එක index අනුව ලබාගන්න පුළුවන්. Python හි counting පටන් ගන්නේ 0 සිට. Row index එක order_id එකට සමාන නොවේ.

WATCH THE STEPS

1	2	3
Index	Read row	Print



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 01_first_pandas.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
print(sales.iloc[0]["product"])
print(sales.iloc[1]["product"])
```

■ Windows Command Prompt / example output

```
Pencil
Notebook
```

REMEMBER

Row index 0 සිට ගණන් කරනවා; order_id වෙනම field එකක්.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

`iloc[2]` හි product එක කුමක්ද?

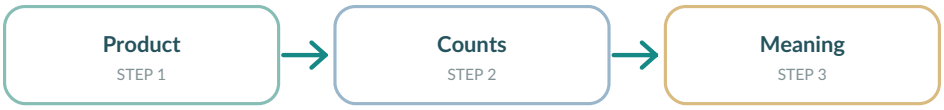
UNIT 04 / LESSON 09

value_counts() - වර්ග ගණන

Product නම් කී වරක් පෙනෙනවාද කියා `value_counts()` මඟින් සොයන්න පුළුවන්. මෙය rows ගණනයි; අලෙවි වූ units ගණන සමඟ එකම දෙයක් නොවේ.

WATCH THE STEPS

1. Group names
2. Count rows
3. Explain



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 01_first_pandas.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
print(sales["product"].value_counts().head())
```

■ Windows Command Prompt / example output

```
Pencil          10
Notebook        10
...
```

REMEMBER

Row count එක sales lines ගණන; units sold වෙනම ගණනයක්.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Product rows ගණන සහ Quantity sum එක වෙනස් වන්නේ ඇයි?

UNIT 04 / LESSON 10

info() සහ describe()

`info()` මඟින් column types සහ non-null counts බලන්න පුළුවන්. `describe()` මඟින් numeric columns සඳහා count, mean, minimum, maximum වැනි සාරාංශ ලබා ගත හැක.

WATCH THE STEPS

1. Inspect metadata
2. Summarize
3. Question



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 01_first_pandas.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
sales.info()
print(sales.describe(include="all"))
```

■ Windows Command Prompt / example output

```
<class pandas.core.frame.DataFrame>
RangeIndex: 61 entries
```

REMEMBER

Raw table එකේ වැරදි rows තිබෙන නිසා statistics පරීක්ෂා කළ පසුව පමණක් විශ්වාස කරන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

describe() හි mean එක අසමාන values නිසා misleading විය හැක්කේ කොහොමද?

UNIT 04 • THINK AND CHOOSE

කුඩා දැනුම් පරීක්ෂාව

වැරදුණොත් පසුගිය පාඩම නැවත කියවන්න. Answers අවසානයේ තිබෙනවා; මුලින්ම ඔබම උත්සාහ කරන්න.

1. pandas table එක?

A. DataFrame
B. Textbox
C. Browser

2. මුල් rows 5 පෙන්වන්නේ?

A. head()
B. tail()
C. copy()

3. shape කියන්නේ?

A. Color
B. Rows සහ columns
C. Price

කඩදාසියේ පිළිතුරු දෙන්න

1. මෙම Unit එකේ ඉගෙනගත් වැදගත්ම අදහස කුමක්ද?
2. මේ සතියේ ඔබම run කළ Python code එක කුමක්ද?
3. ඊළඟට වැඩිදුර පැහැදිලි කරගන්න අවශ්‍ය ප්‍රශ්නය කුමක්ද?

MINI MISSION

මිතුරෙකුට මිනිත්තු දෙකකින් අද පාඩමේ අදහස පැහැදිලි කරන්න.

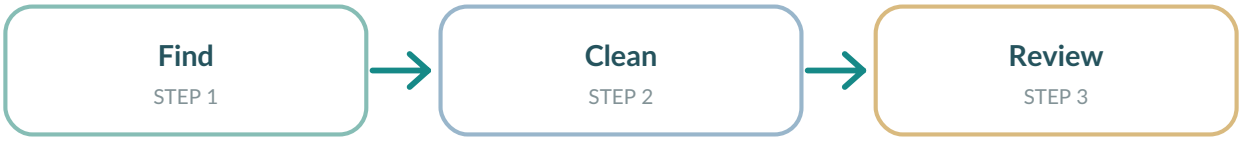
EGOTECHWORLD • DATA ANALYSIS KIDS LAB • 2026

05

වැරදි දත්ත සකස් කරමු

Data Cleaning

Missing, duplicate, invalid dates සහ bad numbers.



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

WHAT YOU WILL LEARN

12 කුඩා පාඩම්, code examples සහ නිවැරදි data reasoning

TRY IT YOURSELF

සෑම පිටුවකම mini mission. Unit අවසානයේ MCQ ත්‍රිත්වයක්.

> දත්ත කියවන්න. සාක්ෂි පරීක්ෂා කරන්න. අවධානයෙන් නිගමනය කරන්න.

UNIT 05 / LESSON 01

Data Cleaning කියන්නේ ඇයි?

Data clean කිරීම කියන්නේ missing, duplicate, invalid වැනි ප්‍රශ්න හඳුනාගෙන යම් පැහැදිලි නීති යටතේ නිවැරදි කිරීම හෝ review සඳහා වෙන් කිරීමයි. අපේ project එකේ raw CSV වෙනස් නොකර clean CSV වෙනම save කරමු.

WATCH THE STEPS

1. Find issues
2. Apply rules
3. Save both



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 02_clean_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
raw = pd.read_csv(SOURCE)
clean, rejected = clean_sales(raw)
print(len(raw), len(clean), len(rejected))
```

■ Windows Command Prompt / example output

```
Raw: 61
Clean: 55
Rejected: 6
```

REMEMBER

Data row එකක් ඉවත් කිරීමෙන් totals වෙනස් වෙයි; ඉවත් කළ හේතුව ලියා තබන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Raw සහ Clean files දෙකම තබාගැනීම වැදගත් ඇයි?

UNIT 05 / LESSON 02

Missing values - හිස් අගයන්

Quantity එක හිස් නම් විකුණූ ගණන දන්නේ නැහැ. එය 1 බවට හිතාගෙන වෙනස් කිරීමෙන් ආදායම වැරදි වෙන්න පුළුවන්. පළමුව pandas `isna()` මඟින් හිස් values හඳුනාගන්නවා.

WATCH THE STEPS

1. Spot missing
2. Count
3. Investigate



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 01_first_pandas.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
print(sales.isna().sum())
missing = sales["quantity"].isna()
print(sales[missing])
```

■ Windows Command Prompt / example output

```
quantity      1
other fields  0
```

REMEMBER

Missing value = නොදන්නා අගය; 0 සමඟ සමාන නැහැ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

හිස් quantity ඇති row එකක් වෙනම සටහන් කරගත යුත්තේ ඇයි?

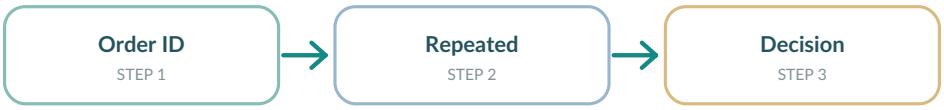
UNIT 05 / LESSON 03

Duplicates - දෙවරක් ලියා ඇති දේ

order_id එක එක වාරයක් පමණක් තිබිය යුතු බවට අපේ sample එකේ නීතියක් තියෙනවා. එකම order_id නැවත ආවොත් revenue දෙවරක් ගණනය විය හැක. `duplicated()` මඟින් හඳුනාගමු.

WATCH THE STEPS

1	2	3
Choose key	Check duplicate	Review



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 02_clean_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
duplicate = df["order_id"].duplicated(keep="first")
print(df.loc[duplicate, ["order_id", "product"]])
```

■ Windows Command Prompt / example output

```
duplicate order_id: S010
Only one copy is kept
```

REMEMBER

"Duplicate" නීතිය dataset design එකට අනුව වෙනස් විය හැක.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

ලෝකයේ සෑම dataset එකකම duplicate IDs ඉවත් කළ යුතුද? හේතුව දක්වන්න.

UNIT 05 / LESSON 04

Spaces සහ Capital letters

`" JUICE "` සහ `"Juice"` එකම product වුණත් text වෙනස්. Leading/trailing spaces ඉවත් කර title-case කරමු. එයින් groupby කරන විට එකම product එකට ගැළපේ.

WATCH THE STEPS

1. Trim spaces
2. Normalize case
3. Compare



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 02_clean_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
df["product"] = (df["product"].astype("string")
                 .str.strip().str.title())
```

■ Windows Command Prompt / example output

```
"  JUICE  "  ->  "Juice"
" stationery " -> "Stationery"
```

REMEMBER

සෑම අකුරු රටාවක්ම title-case කිරීම සැබෑ brand names සඳහා සුදුසු නොවිය හැක.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Product names case-sensitive වුණොත් අලුත් groups වැරදි ලෙස සෑදෙන්නේ කොහොමද?

UNIT 05 / LESSON 05

Numbers convert කිරීම

Quantity column එකේ `two` වැනි text ඇත්නම් `pd.to_numeric(errors="coerce")` මඟින් ගණනය කළ නොහැකි අගය NaN බවට පත් කරන්න පුළුවන්. පසුව අපි ඒ row එක review සඳහා වෙන් කරමු.

WATCH THE STEPS

1. Convert
2. Mark invalid
3. Review



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 02_clean_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
df["quantity"] = pd.to_numeric(
    df["quantity"], errors="coerce")
```

■ Windows Command Prompt / example output

```
"2" -> 2
"two" -> NaN (invalid)
```

REMEMBER

Coerce කළ අගය NaN වූ විට එය මඟහැර report එකක් සෑදෙන්නේ නොදැනීමෙන් විය හැක.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

`two` අගය 2 කරලා replace නොකර review කරන්න කියන්නේ ඇයි?

UNIT 05 / LESSON 06

Dates parse කිරීම

Date එක YYYY-MM-DD ලෙස තිබුණත් `2026-02-30` වැනි නොපවතින දිනයක් ලියන්න පුළුවන්. `pd.to_datetime` මඟින් හරි dates පමණක් date type බවට පත් කරමු.

WATCH THE STEPS

1. Read date
2. Validate
3. Flag invalid



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 02_clean_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
df["date"] = pd.to_datetime(
    df["date"],
    format="%Y-%m-%d", errors="coerce")
```

■ Windows Command Prompt / example output

```
2026-09-01 -> valid date
2026-02-30 -> NaT
```

REMEMBER

NaT = Not a Time. Date එක වැරදි බවට තොරතුරක්.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

සැප්තැම්බර් මාසයේ sales chart එකකට වැරදි February date එකක් ඇතුළත් වුණොත්?

UNIT 05 / LESSON 07

Quantity positive විය යුතුයි

අපේ data model එකේ Quantity යනු ධන පූර්ණ අංකයක් විය යුතුයි. Zero, negative, decimal quantities මේ simple example එකේ වැරදි ලෙස සැලකේ. සත්‍ය ව්‍යාපාරයක refunds වෙනම නීති ඇති.

WATCH THE STEPS

1. Define rule
2. Check quantities
3. Review



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

Visual Studio Code / 02_clean_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
invalid_qty = (df["quantity"] <= 0) | (df["quantity"] % 1 != 0)
print(invalid_qty.sum())
```

Windows Command Prompt / example output

```
-1 -> rejected
0  -> rejected
2  -> accepted
```

REMEMBER

මෙහි නීති sample shop එකට අදාළයි; සියලු industries වලට එකම නීති නොවේ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Return sales දත්ත තිබුණොත් negative quantity සඳහා වෙනම logic එකක් අවශ්‍යද?

UNIT 05 / LESSON 08

Price positive විය යුතුයි

මිල 0 නම් මෙහි revenue analysis සඳහා invalid row ලෙස සලකමු. සැබෑ ලෝකයේ free gifts හෝ discounts සඳහා 0 price වලංගු විය හැක; ඒ නිසා business rules තහවුරු කරගත යුතුයි.

WATCH THE STEPS

1. Read price
2. Apply demo rule
3. Review



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 02_clean_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
df["unit_price"] = pd.to_numeric(
    df["unit_price"], errors="coerce")
invalid_price = df["unit_price"] <= 0
```

■ Windows Command Prompt / example output

```
price = 0 -> rejected in this demo
```

REMEMBER

0 price = වැරදි යන නීතිය අපේ demo එක සඳහා පමණයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Free sample එකක් විකුණුවා කියා සටහන් කළොත් price 0 අවසර දිය යුතුද?

UNIT 05 / LESSON 09

Rejected rows පරීක්ෂා කරමු

වැරදි rows delete කර අමතක කරන්නේ නැහැ. `rejected_rows.csv` නමින් වෙනම save කර අලුතෙන් original source පරීක්ෂා කළ හැක. Real business එකක අහිමි revenue නොවීමට human review වැදගත්.

WATCH THE STEPS

1. Separate
2. Save
3. Investigate



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 02_clean_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
rejected.to_csv(REJECTED, index=False)
print("Review:", len(rejected), "rows")
```

■ Windows Command Prompt / example output

```
Original rows : 61
Rejected rows : 6
Clean rows    : 55
```

REMEMBER

"Rejected" = දැනට report එකට ගත්තේ නැහැ; තොරතුරු සම්පූර්ණයෙන් ඉවත් කළේ නැහැ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Rejected rows 6ක් අතරින් එකකට නිවැරදි අගයක් ලැබුණොත් මීළඟ පියවර කුමක්ද?

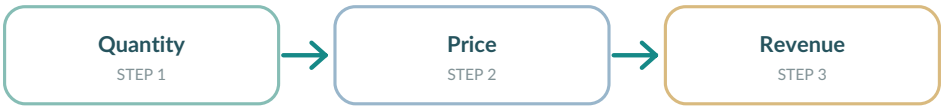
UNIT 05 / LESSON 10

New column: Revenue

Clean row එකේ Quantity සහ Unit Price වලින් Revenue column එකක් හදමු. `df["revenue"]` කියන්නේ DataFrame එකට අලුත් column එකක්. pandas columns මඟින් එකවර rows ගණනාවකට calculation කරන්න පුළුවන්.

WATCH THE STEPS

1. Valid quantity
2. Multiply
3. New column



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 02_clean_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
clean["revenue"] = (
    clean["quantity"] * clean["unit_price"]
)
```

■ Windows Command Prompt / example output

```
Pencil: 2 x 40 = Rs 80
Notebook: 2 x 180 = Rs 360
```

REMEMBER

Revenue column එක ආදායමයි; ණය, වියදම්, බදු නැතිව profit කියන්න එපා.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Quantity 3 සහ Unit Price 90 නම් Revenue කීයද?

UNIT 05 / LESSON 11

Clean file save කරන්න

`to_csv()` මඟින් cleaned DataFrame එක අලුත් CSV file එකක් බවට පත් කරනවා.
`index=False` කියන්නේ අමතර pandas row index column එකක් CSV තුළ නොලියන ලෙසයි.

WATCH THE STEPS

1. Clean
2. Export
3. Reopen



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 02_clean_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
clean.to_csv("outputs/shop_sales_clean.csv", index=False)
print(len(clean))
```

■ Windows Command Prompt / example output

```
Clean rows    : 55
Saved: shop_sales_clean.csv
```

REMEMBER

Raw data වෙනම තබා නැවත cleaning process එක repeat කළ හැකි විය යුතුයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

`index=False` නැතිනම් output CSV තුළ වෙනස් වන දෙයක් කියන්න.

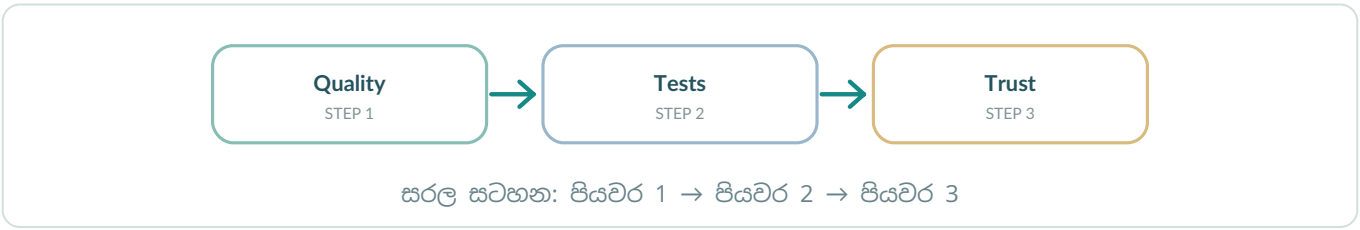
UNIT 05 / LESSON 12

Quality checklist එක

Report එක ගන්න කලින් rows, duplicate IDs, missing fields, dates සහ positive quantity/price හරිද බලමු. වැරදි source දත්ත නිසා වැරදි business conclusions ගත හැකි නිසා මෙය අනිවාර්ය පියවරකි.

WATCH THE STEPS

1. Check keys
2. Check values
3. Check totals



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / tests/test_analysis.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
assert not clean["order_id"].duplicated().any()
assert (clean["quantity"] > 0).all()
assert (clean["unit_price"] > 0).all()
```

■ Windows Command Prompt / example output

```
5 tests passed (sample project)
```

REMEMBER

Assertions පසුව code changes වල වැරදි අනාවරණය කිරීමටත් උදව් වේ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Check items තුනක් ඔබගේ notebook එකේ checkbox ලෙස ලියන්න.

UNIT 05 • THINK AND CHOOSE

කුඩා දැනුම් පරීක්ෂාව

වැරදුණොත් පසුගිය පාඩම නැවත කියවන්න. Answers අවසානයේ තිබෙනවා; මුලින්ම ඔබම උත්සාහ කරන්න.

1. NaN වැනි අගයක්?

A. Missing/invalid value
B. Positive number
C. Chart

2. Duplicate order id හොඳද?

A. Always
B. Check before counting
C. Never check

3. rejected_rows.csv එක?

A. Human review සඳහා
B. Share personal data
C. Web app

කඩදාසියේ පිළිතුරු දෙන්න

1. මෙම Unit එකේ ඉගෙනගත් වැදගත්ම අදහස කුමක්ද?
2. මේ සතියේ ඔබම run කළ Python code එක කුමක්ද?
3. ඊළඟට වැඩිදුර පැහැදිලි කරගන්න අවශ්‍ය ප්‍රශ්නය කුමක්ද?

MINI MISSION

මිතුරෙකුට මිනිත්තු දෙකකින් අද පාඩමේ අදහස පැහැදිලි කරන්න.

EGOTECHWORLD • DATA ANALYSIS KIDS LAB • 2026

06

ගණන් සහ groupby

Simple Statistics

Revenue, units, average, categories, daily summaries.



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

WHAT YOU WILL LEARN

10 කුඩා පාඩම්, code examples සහ නිවැරදි data reasoning

TRY IT YOURSELF

සෑම පිටුවකම mini mission. Unit අවසානයේ MCQ ත්‍රිත්වයක්.

> දත්ත කියවන්න. සාක්ෂි පරීක්ෂා කරන්න. අවධානයෙන් නිගමනය කරන්න.

UNIT 06 / LESSON 01

Total Revenue - මුළු අගය

Clean data වල revenue column එකේ අගයන් එකතු කළ විට Rs 20,120 ලැබෙනවා. මෙය sample data එකේ sales line 55ක් සඳහාය. බදු, discounts හෝ product costs මෙහි සටහන් නොවේ.

WATCH THE STEPS

1. Clean rows
2. Add revenue
3. Label total



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 03_analyze_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
total_revenue = clean["revenue"].sum()
print(f"Rs {total_revenue:,.2f}")
```

■ Windows Command Prompt / example output

```
Revenue (not profit): Rs 20,120.00
```

REMEMBER

Total අගය sample date range එකට පමණක් අදාළයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Revenue කියන metric එක පිළිබඳ report එකක සීමාව කොහොමද ලියන්නේ?

UNIT 06 / LESSON 02

Units Sold - භාණ්ඩ ගණන

Sales line ගණන සහ units ගණන දෙක වෙනස්. එක් row එකක quantity 4 විය හැක. අපේ clean rows 55 තුළ සමස්ත units 135ක් විකුණා තිබේ.

WATCH THE STEPS

1. Read quantities
2. Sum
3. Compare



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 03_analyze_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
total_units = clean["quantity"].sum()
print("Units sold:", total_units)
```

■ Windows Command Prompt / example output

```
Valid sale lines : 55
Units sold       : 135
```

REMEMBER

Transaction count ලෙස order_id unique බවත් සලකා බලන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Sale lines 55 සහ units 135 අතර වෙනස පෙළ දෙකකින් කියන්න.

UNIT 06 / LESSON 03

Average Revenue Per Line

Revenue එක sale lines ගණනෙන් බෙදුවහම එක line එකකට සාමාන්‍ය ආදායම ලැබේ. Clean data වල එය Rs 365.82 (වට කළ අගය) පමණයි. මෙය එක් customer කෙනෙකුගේ average spend නොවේ.

WATCH THE STEPS

1. Revenue
2. Divide by lines
3. Interpret



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 03_analyze_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
mean_line = clean["revenue"].mean()
print(f"Rs {mean_line:,.2f}")
```

■ Windows Command Prompt / example output

```
Mean revenue / line: Rs 365.82
```

REMEMBER

එක් row එක = එක් sale line එකක්. Customer-level ID මේ dataset එකේ නැහැ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Average line revenue එක customer average ලෙස නොලියන්නේ ඇයි?

UNIT 06 / LESSON 04

Min, Max සහ Median

Minimum සහ maximum කුඩාම හා විශාලම values පෙන්වයි. Median කියන්නේ values පිළිවෙළට තැබුවිට මැද අගය. Extreme values වලින් mean එකට බලපෑම තිබුණත් median වඩා ස්ථාවර විය හැක.

WATCH THE STEPS

1. Sort values
2. Find center
3. Compare



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 03_analyze_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
print(clean["revenue"].min())
print(clean["revenue"].max())
print(clean["revenue"].median())
```

■ Windows Command Prompt / example output

```
min / max / median = summary statistics
```

REMEMBER

Summary statistics තනිවම context නොදෙයි; distribution එකත් බලන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

[10,20,30,1000] සඳහා median සහ mean වෙනස ගණනය කරන්න.

UNIT 06 / LESSON 05

groupby() - කොටස් කිරීම

`groupby("category")` කියන්නේ එකම category ඇති rows එකට ගන්නා ක්‍රමය. එම කණ්ඩායම තුළ Revenue sum කළාම category එකකට මුළු මුදල ලැබේ.

WATCH THE STEPS

1	2	3
Split groups	Sum each	Combine



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 03_analyze_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
by_category = clean.groupby("category")["revenue"].sum()
print(by_category)
```

■ Windows Command Prompt / example output

```
Accessories   10500
Snacks         4620
Stationery     5000
```

REMEMBER

Groupby සංඛ්‍යා එකතු කරනවා; business හේතුවක් ස්වයංක්‍රීයව සොයාදෙන්නේ නැහැ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Category තුනෙන් වැඩිම revenue ඇත්තේ කුමකටද?

UNIT 06 / LESSON 06

agg() - ගණන් දෙකක් එකවර

`agg()` මඟින් එකම group සඳහා revenue සහ units එකවර ගන්න පුළුවන්. එවිට එක category එකක විකුණුම් මුදල සහ භාණ්ඩ ගණන අතර වෙනස පෙනෙයි.

WATCH THE STEPS

1	2	3
Group	Aggregate	Compare



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 03_analyze_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
table = clean.groupby("category", as_index=False).agg(
    revenue=("revenue", "sum"),
    units=("quantity", "sum"))
```

■ Windows Command Prompt / example output

```
Accessories : revenue 10500, units 30
Stationery  : revenue  5000, units 60
```

REMEMBER

Units සහ Revenue අතර unit price එක විශාල බලපෑමක් දක්වයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Stationery units 60ක් වුවත් Accessories revenue වැඩි වීමට හේතුවක් යෝජනා කරන්න.

UNIT 06 / LESSON 07

sort_values() - පිළිවෙළට

Highest revenue product එක මුලට ගන්න `sort_values(ascending=False)` භාවිතා කරමු. Descending කියන්නේ ඉහළ අගයේ සිට පහළට.

WATCH THE STEPS

1. Group
2. Sort
3. Read top



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 03_analyze_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
top = (clean.groupby("product")["revenue"].sum()
       .sort_values(ascending=False))
print(top.head(3))
```

■ Windows Command Prompt / example output

```
Top product: Water Bottle
```

REMEMBER

Top revenue product සහ most units sold product වෙනස් විය හැක.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Ascending=True දැම්මොත් පළමුව පෙනෙන්නේ කුමන වර්ගයේ items ද?

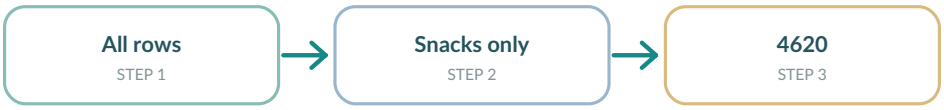
UNIT 06 / LESSON 08

Filter - අවශ්‍ය rows පමණක්

`clean[clean["category"] == "Snacks"]` මඟින් Snacks category එකට අයත් rows පමණක් තෝරන්න පුළුවන්. Filter එකක් යෙදූ පසු whole-shop total එක සමඟ ගැටලුවක් නොවීමට පැහැදිලිව label කරන්න.

WATCH THE STEPS

1. Choose category
2. Filter rows
3. Sum



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 03_analyze_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
snacks = clean[clean["category"] == "Snacks"]
print(snacks["revenue"].sum())
```

■ Windows Command Prompt / example output

```
Snacks revenue: Rs 4,620
```

REMEMBER

Filter නිසා report scope එක වෙනස් වන බව chart title එකේ සඳහන් කරන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

`Stationery` category එක පමණක් filter කරන code එක ලියන්න.

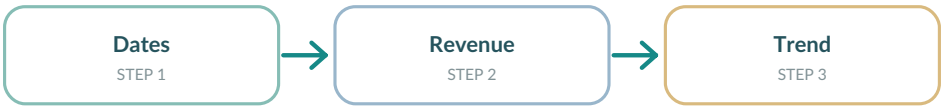
UNIT 06 / LESSON 09

Daily groupby - දිනකට විකුණුම්

`groupby("date")` මඟින් එක් දිනක සියලු revenue values එකතු කර daily total එක සොයාගන්න පුළුවන්. Dates නිවැරදිව sort කළ විට කාලයත් සමඟ වෙනස්වීම chart එකෙන් පෙනෙයි.

WATCH THE STEPS

1. Group date
2. Add totals
3. Sort dates



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 03_analyze_sales.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
daily = clean.groupby("date", as_index=False).agg(
    revenue=("revenue", "sum"))
daily = daily.sort_values("date")
```

■ Windows Command Prompt / example output

```
date         revenue
2026-09-01   ...
2026-09-02   ...
```

REMEMBER

Daily variation එකක් තනිවම growth trend එකක් බව ඔප්පු නොකරයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

දින දෙකක sales අඩු වුණොත් හේතුව තීරණය කිරීමට තවත් දත්ත අවශ්‍යද?

UNIT 06 / LESSON 10

Group totals සහ grand total සමානද?

Category totals එකතු කළ විට මුළු revenue එකට සමාන විය යුතුයි. අපේ sample එකේ 10,500 + 5,000 + 4,620 = 20,120. මෙය data sanity check එකක්.

WATCH THE STEPS

1. Calculate totals
2. Compare
3. Assert



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / tests/test_analysis.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
category_total = table["revenue"].sum()
overall_total = clean["revenue"].sum()
assert category_total == overall_total
```

■ Windows Command Prompt / example output

```
Category total = Overall total = 20120
```

REMEMBER

Grouping key වල missing values තිබුණොත් sum comparisons විශේෂයෙන් බලන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Category එකක් missing වුණොත් දෙක සමාන නොවිය හැක්කේ කොහොමද?

UNIT 06 • THINK AND CHOOSE

කුඩා දැනුම් පරීක්ෂාව

වැරදුණොත් පසුගිය පාඩම නැවත කියවන්න. Answers අවසානයේ තිබෙනවා; මුලින්ම ඔබම උත්සාහ කරන්න.

1. Groupby එකේ වැඩ?

A. Groups එකතු කිරීම
B. Image draw
C. Install software

2. Mean කියන්නේ?

A. Sum / count
B. Maximum
C. Minimum

3. Revenue 20,120 කියන්නේ profit ද?

A. Yes
B. No
C. Maybe always

කඩදාසියේ පිළිතුරු දෙන්න

1. මෙම Unit එකේ ඉගෙනගත් වැදගත්ම අදහස කුමක්ද?
2. මේ සතියේ ඔබම run කළ Python code එක කුමක්ද?
3. ඊළඟට වැඩිදුර පැහැදිලි කරගන්න අවශ්‍ය ප්‍රශ්නය කුමක්ද?

MINI MISSION

මිතුරෙකුට මිනිත්තු දෙකකින් අද පාඩමේ අදහස පැහැදිලි කරන්න.

EGOTECHWORLD • DATA ANALYSIS KIDS LAB • 2026

07

Charts හදා කියවමු

Charts and Stories

Bar charts, line charts, units සහ නිවැරදි visual stories.



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

WHAT YOU WILL LEARN

9 කුඩා පාඩම්, code examples සහ නිවැරදි data reasoning

TRY IT YOURSELF

සෑම පිටුවකම mini mission. Unit අවසානයේ MCQ ත්‍රිත්වයක්.

> දත්ත කියවන්න. සාක්ෂි පරීක්ෂා කරන්න. අවධානයෙන් නිගමනය කරන්න.

UNIT 07 / LESSON 01

Graph කියන්නේ කතාවක්

Graph එකක් මඟින් numbers පහසුවෙන් සසඳා බලන්න පුළුවන්. නමුත් graph එකක් ලස්සන වුණාට ප්‍රශ්නයකට හොඳ පිළිතුරක් දිය යුතුයි. ඉතා ලොකු අගයන් visually misleading නොවීමට axes සහ labels දමන්න.

WATCH THE STEPS

1. Choose question
2. Choose plot
3. Label



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 04_make_charts.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
import matplotlib.pyplot as plt
print("Ready to draw charts")
```

■ Windows Command Prompt / example output

```
Saved: charts/01_categories.png
```

REMEMBER

Chart එකේ source “Fictional sample data” ලෙස සඳහන් කරන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Chart එකකට title, labels, units අවශ්‍ය ඇයි?

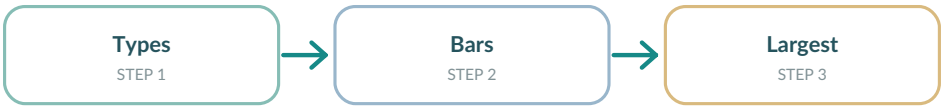
UNIT 07 / LESSON 02

Bar chart - categories සසඳමු

Bar chart එක category වගේ වෙන් වෙන් වර්ගවල අගයන් සසඳන්න හොඳයි. Accessories, Stationery සහ Snacks revenue අපි ඉතා පැහැදිලි තීරු තුනකින් පෙන්වමු.

WATCH THE STEPS

1	2	3
Categories	Bar heights	Compare



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 04_make_charts.py

EXPLORER
data-analysis/
run_all.py
data/
shop_sales_raw.csv
lessons/
00_hello.py
outputs/
charts/
tests/

```
fig, ax = plt.subplots()
ax.bar(by_category["category"], by_category["revenue"])
ax.set_ylabel("Revenue (LKR)")
```

■ Windows Command Prompt / example output

```
Accessories  | ########## 10500
Stationery   | #####       5000
Snacks       | ####        4620
```

REMEMBER

Bar height ඉහළ යන බව = price වැඩි / units වැඩි යන දෙකෙන් එකක් හෝ දෙකම විය හැක.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Bar chart එකේ වැඩිම revenue ඇති category එක හඳුනාගන්න.

UNIT 07 / LESSON 03

Line chart - දින අනුව

Time හරහා value එකක් වෙනස් වන හැටි දැක්වීමට line chart එක හොඳයි. Day-to-day revenue සඳහා points සහ lines අඳිමු. Date order වැරදි නම් chart එක තේරුම් ගැනීමට අමාරු වේ.

WATCH THE STEPS

1. Sort dates
2. Plot points
3. Connect



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 04_make_charts.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
ax.plot(by_day["date"], by_day["revenue"], marker="o")
ax.set_title("Daily Revenue - Sample Data")
```

■ Windows Command Prompt / example output

```
Day 01 - total
Day 02 - total
Day 03 - total
... plotted over time
```

REMEMBER

Line chart එකෙන් future sales predict වෙන්නේ නැහැ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Daily line chart එක බැලූ පසු දත්තවල උච්ච අගයක් තිබෙන දවසක් සොයාගන්න.

UNIT 07 / LESSON 04

Axes - x සහ y

Horizontal axis = x-axis; vertical axis = y-axis. Bar chart එකේ x-axis category; y-axis Revenue (LKR). Unit එක නැතිව 10,500 කියන අගය රුපියල්ද pieces ද කියන්න බැහැ.

WATCH THE STEPS

1. Name x
2. Name y
3. Add unit



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 04_make_charts.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
ax.set_xlabel("Category")
ax.set_ylabel("Revenue (LKR)")
ax.set_title("Sales by Category")
```

■ Windows Command Prompt / example output

```
X: Category
Y: Revenue (LKR)
```

REMEMBER

Axes වල units පැහැදිලිව පෙන්වීම chart quality එක ඉහළ නංවයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

යම් chart එකක y-axis “Sales” විතරක් දැක්වුවොත් තව කුමක් පැහැදිලි කළ යුතුද?

UNIT 07 / LESSON 05

Plot save කිරීම

`plt.show()` මඟින් computer screen එකේ graph එකක් දකින්න පුළුවන්. අපේ project එකේ `fig.savefig()` භාවිතයෙන් `.png` file එකක් save කරනවා, ඒ නිසා headless CMD මඟිනුත් වැඩ කරයි.

WATCH THE STEPS

1	2	3
Figure	savefig	PNG



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 04_make_charts.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
fig.tight_layout()
fig.savefig("charts/01_categories.png", dpi=160)
plt.close(fig)
```

■ Windows Command Prompt / example output

```
Charts saved in: .../charts
```

REMEMBER

Draw | Save PNG | Open file

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Chart file name එකෙන් content එක කියන්න පුළුවන් ලෙස නම් කරන්න.

UNIT 07 / LESSON 06

Bar chart එකේ සත්‍ය කතාව

Accessories revenue Rs 10,500; Stationery Rs 5,000; Snacks Rs 4,620. මේ chart එකෙන් sample එකේ Accessories වැඩි ආදායම ලබාදුන් බව පමණක් පෙන්වයි. Stock decisions සඳහා තවත් තොරතුරු අවශ්‍යයි.

WATCH THE STEPS

1. Look
2. Compare
3. Explain

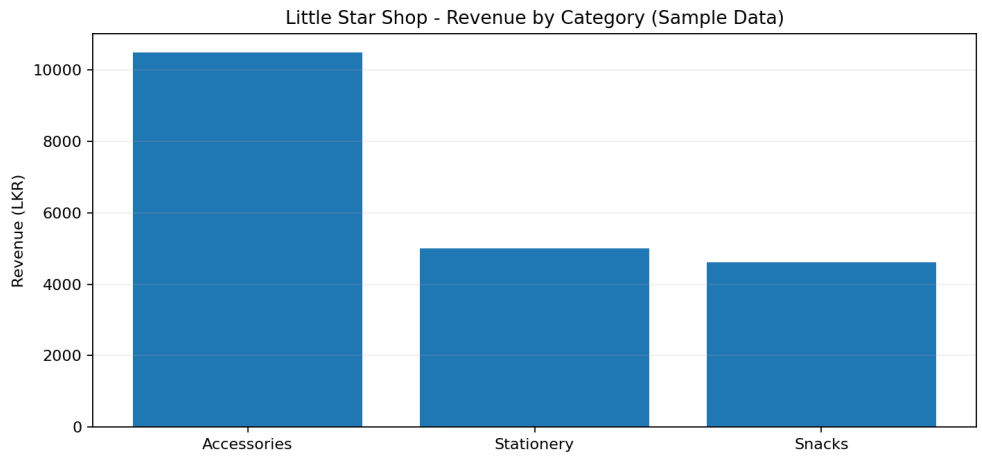


CHART SOURCE

charts/01_categories.png • generated from the cleaned CSV files

REMEMBER

අපේ graph එක software-generated preview එකක්; සැබෑ shop chart එකක් නොවේ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Accessories units අඩු වුණත් revenue ඉහළට යන හැකියාව ගැන හේතුවක් ලියන්න.

UNIT 07 / LESSON 07

Line chart එකේ සීමා

Daily trend chart එකේ සමහර දවස්වල revenue ඉහළටත් පහළටත් යනවා. මේ fluctuations එකෙන් අනාගතය “නියතව” පෙන්වන්න බැහැ. කාලය, promotion, stock availability වැනි දේවල් අදාළ වේ.

WATCH THE STEPS

1. Observe daily
2. Notice changes
3. Ask why

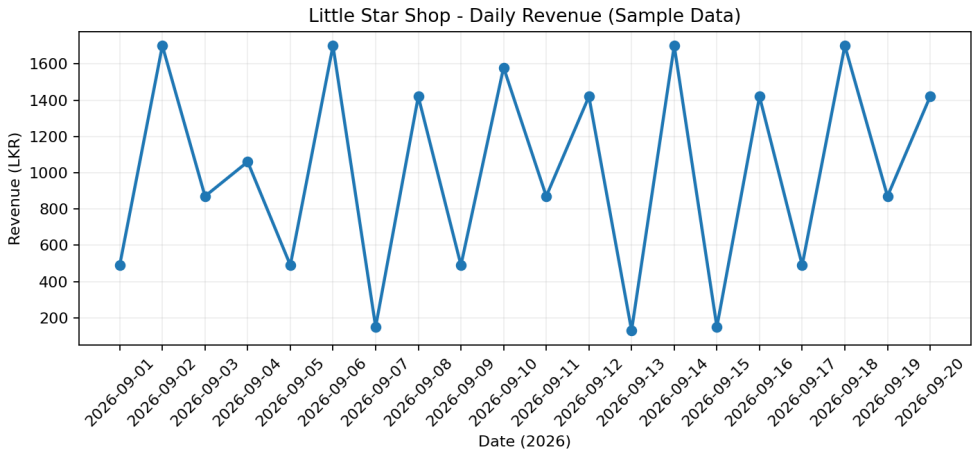


CHART SOURCE

charts/02_daily_trend.png • generated from the cleaned CSV files

REMEMBER

Sample date range කුඩා නිසා future trend ගැන ශක්තිමත් නිගමනයක් ගන්න බැහැ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Line chart එකේ change එකක් ඇත්නම් ඒ සඳහා නිශ්චිත හේතුව නොදන්නා බව ලියන්න.

UNIT 07 / LESSON 08

Misleading graphs - අනතුරු

Y-axis එක 9000 සිට පටන් ගත්තොත් 10,500 සහ 5,000 අතර වෙනස වඩා විශාල ලෙස පෙනෙනවා. Bar charts වල zero baseline එක සාමාන්‍යයෙන් හොඳයි; chart එය වෙනස් කරනවා නම් පැහැදිලිව සඳහන් කරන්න.

WATCH THE STEPS

1	2	3
Read title	Inspect scale	Verify



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 04_make_charts.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
ax.set_ylim(bottom=0)
ax.grid(axis="y", alpha=0.2)
```

■ Windows Command Prompt / example output

```
Good bar chart: y starts at 0
Danger: cropped y-axis without notice
```

REMEMBER

විශේෂයෙන් percentages, units, zero baseline පරීක්ෂා කරන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Chart එකක axes නොපෙනෙනවා නම් තේරුම් ගැනීමට ඇති අපහසුතාව කියන්න.

UNIT 07 / LESSON 09

Chart එකකට නිවැරදි වාක්‍යයක්

Charts ගැන ලිවීමත් ඉගෙනුමේ කොටසක්. “Sample records වල Accessories category එක වැඩිම revenue ලබාදුන්නා” හොඳයි. “Accessories ලෝකයේ සෑම කඩයකම හොඳයි” කියන නිගමනය වැරදියි.

WATCH THE STEPS

1. Read
2. State finding
3. Limit scope



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / notes/chart_insight.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
insight = ("In this fictional sample, "
           "Accessories had the highest revenue.")
print(insight)
```

■ Windows Command Prompt / example output

```
One chart -> One grounded sentence
```

REMEMBER

Insight යනු evidence + context + limitations.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Graph එකක් බැලූ පසු “these sample rows” යන වචන යොදා නිගමනයක් ලියන්න.

UNIT 07 • THINK AND CHOOSE

කුඩා දැනුම් පරීක්ෂාව

වැරදුණොත් පසුගිය පාඩම නැවත කියවන්න. Answers අවසානයේ තිබෙනවා; මුලින්ම ඔබම උත්සාහ කරන්න.

1. Categories compare කිරීම?

A. Bar chart
B. Audio
C. Text file

2. Time trend පෙන්වන්නේ?

A. Line chart
B. Folder list
C. Column name

3. Chart labels?

A. Useful
B. Not needed
C. Always secret

කඩදාසියේ පිළිතුරු දෙන්න

1. මෙම Unit එකේ ඉගෙනගත් වැදගත්ම අදහස කුමක්ද?
2. මේ සතියේ ඔබම run කළ Python code එක කුමක්ද?
3. ඊළඟට වැඩිදුර පැහැදිලි කරගන්න අවශ්‍ය ප්‍රශ්නය කුමක්ද?

MINI MISSION

මිතුරෙකුට මිනිත්තු දෙකකින් අද පාඩමේ අදහස පැහැදිලි කරන්න.

EGOTECHWORLD • DATA ANALYSIS KIDS LAB • 2026

08

Reports සහ checks

Reports and Tests

Data report, automation, tests සහ conclusions.



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

WHAT YOU WILL LEARN

8 කුඩා පාඩම්, code examples සහ නිවැරදි data reasoning

TRY IT YOURSELF

සෑම පිටුවකම mini mission. Unit අවසානයේ MCQ ත්‍රිත්වයක්.

> දත්ත කියවන්න. සාක්ෂි පරීක්ෂා කරන්න. අවධානයෙන් නිගමනය කරන්න.

UNIT 08 / LESSON 01

Report එකක් කියන්නේ මොකක්ද?

Report එකකින් "මොන දත්තද? මොකක්ද හොයාගත්තේ? සීමාව මොකක්ද?" කියලා පැහැදිලි කරමු. අපි outputs/shop_report.txt තුළ plain-text summary එකක් හදනවා.

WATCH THE STEPS

1. State question
2. Show metrics
3. Mention limits



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 05_final_report.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
report = ["Little Star Shop Report",
          "Revenue: Rs 20,120 (not profit)"]
print("\n".join(report))
```

■ Windows Command Prompt / example output

```
Valid sale lines: 55
Units sold: 135
Total revenue: Rs 20,120.00
```

REMEMBER

Report එකක numbers සහ source යන දෙකම පෙන්වන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Report title එකේ 2026 සහ "Sample Data" දාන්න හේතුව මොකක්ද?

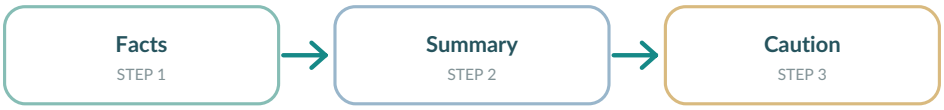
UNIT 08 / LESSON 02

Executive summary - වාක්‍ය තුන

Manager කෙනෙකුට කාලය අඩුයි. ඒ නිසා valid sale lines 55, revenue Rs 20,120, top product Water Bottle යන කරුණු කෙටියෙන් පැහැදිලි කරමු. මේවා example data නිසා සැබෑ ලෝකයට generalize නොකරමු.

WATCH THE STEPS

1. Pick main facts
2. Write concise
3. Check truth



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 05_final_report.py

EXPLORER
data-analysis/
run_all.py
data/
shop_sales_raw.csv
lessons/
00_hello.py
outputs/
charts/
tests/

```
summary = ("55 valid lines; revenue Rs 20,120; "
           "top revenue product: Water Bottle")
print(summary)
```

■ Windows Command Prompt / example output

```
55 clean sale lines
Revenue: Rs 20,120
Top: Water Bottle
```

REMEMBER

ක්‍රියාමාර්ගයක් යෝජනා කළත් එය test කිරීමට වෙනම සැලැස්මක් අවශ්‍යයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Summary එකට සීමාවක් සඳහන් කරන තවත් වාක්‍යයක් එක්කරන්න.

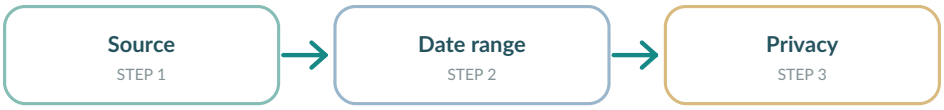
UNIT 08 / LESSON 03

Data source එක ලියමු

Source නැති report එකක් කියවන්නාට verify කරන්න අමාරුයි. අපේ report එකේ CSV file නම, dataset එක fictional බව, row count සහ cleaning rules සඳහන් කරනවා.

WATCH THE STEPS

1. Name file
2. State scope
3. Cite process



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 05_final_report.py

EXPLORER
data-analysis/
run_all.py
data/
shop_sales_raw.csv
lessons/
00_hello.py
outputs/
charts/
tests/

```
source = "data/shop_sales_raw.csv"
print("Source:", source)
print("All records are fictional")
```

■ Windows Command Prompt / example output

```
Source: sample shop CSV
No real customer records
```

REMEMBER

Customer names, phone numbers, payment details tutorial example වල භාවිතා නොකරමු.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Report එකකට source date range සඳහන් කළ යුත්තේ ඇයි?

UNIT 08 / LESSON 04

Report save කරනවා

Python සමඟ report text file එකක් save කළොත් වෙන කෙනෙකුට ඉතා පහසුවෙන් කියවන්න පුළුවන්. අපේ `05_final_report.py` output file එක UTF-8 encoding භාවිතා කරයි.

WATCH THE STEPS

1. Compose
2. Save
3. Open



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / 05_final_report.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
output = ROOT / "outputs" / "shop_report.txt"
output.write_text("Shop summary\n", encoding="utf-8")
```

■ Windows Command Prompt / example output

```
Saved: outputs/shop_report.txt
```

REMEMBER

Actual project එකේ සම්පූර්ණ metrics සමඟ report content සකස් වෙයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

`write_text()` සහ `read_text()` අතර වෙනස පැහැදිලි කරන්න.

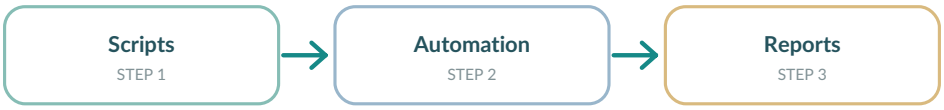
UNIT 08 / LESSON 05

Automate - run_all.py

එකම commands හයක් එකින් එක manually ධාවනය කරනවාට වඩා `run_all.py` ඒවා පිළිවෙළින් ක්‍රියාත්මක කරයි. අලුත් sample data එකක් generate කළ සෑම වරකම එකම steps යොදාගන්න ලේසි.

WATCH THE STEPS

1. List steps
2. Run scripts
3. Collect outputs



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / run_all.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
for file_name in FILES:
    subprocess.run([sys.executable, file_name], check=True)
```

■ Windows Command Prompt / example output

```
C:\> python run_all.py
Done! Open charts/ and outputs/.
```

REMEMBER

Automation එකේ benefit = repeatable process; still verify outputs.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Auto script එක නවත්වන error එකක් එනවිට කුමන file එකෙන් ද කියා සොයාගන්න.

UNIT 08 / LESSON 06

Unit tests - අපේ අගයන් check

අපේ project එකේ unittest tests පහක් තියෙනවා. Raw rows 61, rejected 6, clean 55, duplicates නැති බව, totals එකිනෙකට සමාන බව පරීක්ෂා කරනවා.

WATCH THE STEPS

1. Run tests
2. Read failures
3. Fix



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / CMD • tests

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
python -m unittest discover -s tests -v
```

■ Windows Command Prompt / example output

```
Ran 5 tests
OK
```

REMEMBER

Tests pass වුණා කියලා සෑම business case එකක්ම නිවැරදි බවට සහතිකයක් නොවේ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Quantity සහ unit_price positive බව පරීක්ෂා කිරීමට test එකක් හිතන්න.

UNIT 08 / LESSON 07

Data analyst කෙනෙකුගේ checklist

Final report එක යවන්න කලින් source, missing data, totals, chart labels, units, privacy, සහ conclusions check කරන්න. Coding හොඳ වුණත් interpretation වැරදි නම් වාර්තාව වැරදි යන්න පුළුවන්.

WATCH THE STEPS

1. Check data
2. Check graphics
3. Check words



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / notes/checklist.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
checks = ["source", "cleaning", "totals",
          "labels", "privacy", "limitations"]
for item in checks:
    print("Check:", item)
```

■ Windows Command Prompt / example output

```
CHECK: Source / Cleaning / Totals
CHECK: Labels / Privacy / Limits
```

REMEMBER

Software output එක decision එක නොවේ; human review වැදගත්.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

මිතුරෙකුගේ report එක review කිරීමට ඔබේම checklist එකක් හදන්න.

UNIT 08 / LESSON 08

Mini dashboard එකක් ගැන හිතමු

Dashboard එක කියන්නේ ඉතා වැදගත් metrics සහ charts එකම තැන පෙන්වන screen එකක්. අපේ project එක PNG සහ CSV outputs පදනම් කරගෙන පසුව dashboard එකක් හදන්න පුළුවන්.

WATCH THE STEPS

1. Metrics
2. Charts
3. Explanation



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE

◆ Visual Studio Code / notes/dashboard.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
metrics = {"revenue": 20120, "lines": 55, "units": 135}
print(metrics)
```

■ Windows Command Prompt / example output

```
Revenue    : Rs 20,120
Sale lines: 55
Units      : 135
```

REMEMBER

මෙම PDF හි dashboard එක visual sketch එකක්; අමතර web app එකක් build කරන්නේ නැහැ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Dashboard එකේ card තුනක්, chart දෙකක් ඇඳලා ලේබල් කරන්න.

UNIT 08 • THINK AND CHOOSE

කුඩා දැනුම් පරීක්ෂාව

වැරදුණොත් පසුගිය පාඩම නැවත කියවන්න. Answers අවසානයේ තිබෙනවා; මුලින්ම ඔබම උත්සාහ කරන්න.

1. Report එකට source?

A. අවශ්‍යයි
B. ඕන නැහැ
C. Delete කරන්න

2. Tests 5 pass වුණා කියන්නේ?

A. Everything perfect
B. Checks 5 pass
C. No need review

3. Run_all.py?

A. Commands automate
B. Change prices
C. Creates AI model

කඩදාසියේ පිළිතුරු දෙන්න

1. මෙම Unit එකේ ඉගෙනගත් වැදගත්ම අදහස කුමක්ද?
2. මේ සතියේ ඔබම run කළ Python code එක කුමක්ද?
3. ඊළඟට වැඩිදුර පැහැදිලි කරගන්න අවශ්‍ය ප්‍රශ්නය කුමක්ද?

MINI MISSION

මිතුරෙකුට මිනිත්තු දෙකකින් අද පාඩමේ අදහස පැහැදිලි කරන්න.

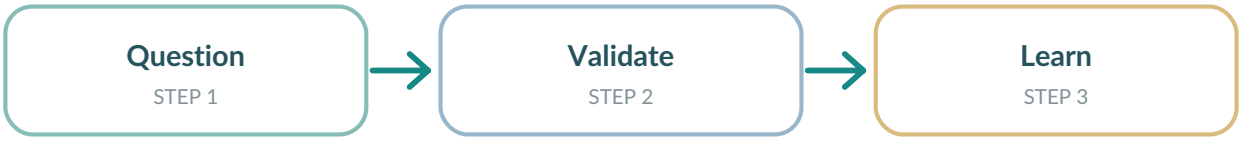
EGOTECHWORLD • DATA ANALYSIS KIDS LAB • 2026

09

AI කාලයේ Data Thinking

2026 Data Literacy

Privacy, bias, AI verification, data ethics සහ next project.



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

WHAT YOU WILL LEARN

8 කුඩා පාඩම්, code examples සහ නිවැරදි data reasoning

TRY IT YOURSELF

සෑම පිටුවකම mini mission. Unit අවසානයේ MCQ ත්‍රිත්වයක්.

> දත්ත කියවන්න. සාක්ෂි පරීක්ෂා කරන්න. අවධානයෙන් නිගමනය කරන්න.

UNIT 09 / LESSON 01

Data Analysis සහ Machine Learning

Data Analysis එකකින් දැනටමත් ඇති data විස්තර කරන්න පුළුවන්. Machine Learning model එකක් training examples වලින් pattern ඉගෙනගෙන අලුත් values predict කිරීමට උත්සාහ කරයි. මේ පොතේ project එක Data Analysis එකක්, ML prediction model එකක් නොවේ.

WATCH THE STEPS

1. Describe past
2. Learn pattern
3. Predict?



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / notes/analysis_vs_ml.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
total = clean["revenue"].sum()
print("Observed revenue:", total)
# This is not a forecast.
```

■ Windows Command Prompt / example output

```
Observed total: Rs 20,120
Not a future prediction
```

REMEMBER

AI, ML, Statistics සහ Data Analysis එකිනෙකට සම්බන්ධ නමුත් එකම දේ නොවේ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Past sales summary එකක් සහ future prediction එකක් වෙන්කර ලියන්න.

UNIT 09 / LESSON 02

AI tools 2026 දී භාවිතා කරන හැටි

AI assistant එකකට pandas code එකක් explain කරන්න හෝ error එකකට අදහසක් යෝජනා කරන්න පුළුවන්. නමුත් AI කියන totals ඇත්ත dataset එකට ගැළපෙනවාද කියා Python code සහ tests වලින් verify කරන්න.

WATCH THE STEPS

1. Ask
2. Verify
3. Explain limits



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / notes/ai_checks.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
# Ask an AI tool to explain groupby, but verify:
assert 10500 + 5000 + 4620 == 20120
```

■ Windows Command Prompt / example output

```
CHECK: 10500 + 5000 + 4620 = 20120
```

REMEMBER

Private customer data free AI chat tools වලට paste කිරීමෙන් වළකින්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

AI assistant එකක් වැරදි revenue number එකක් දුන්නොත් පරීක්ෂා කරන ක්‍රම දෙකක් කියන්න.

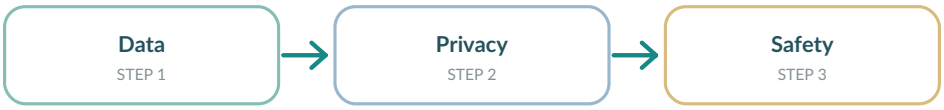
UNIT 09 / LESSON 03

Privacy සහ customer data

සැබෑ customer data වල phone numbers, names, addresses, payment details වැනි දේවල් තිබිය හැක. අපේ පාඩම් සඳහා synthetic example data පමණක් යොදා ගැනීමට හේතුව learners සහ customers රැකගැනීමයි.

WATCH THE STEPS

1. Identify sensitive
2. Minimize
3. Protect



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / notes/privacy.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
row = {"order_id": "S001", "product": "Pencil"}
# Do not include names or payment details.
print(row)
```

■ Windows Command Prompt / example output

```
S001, Pencil
No personal customer fields
```

REMEMBER

Data use කිරීමට පෙර අයිතිය, අවසර සහ privacy පරීක්ෂා කළ යුතුයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

රහස්‍ය දත්ත හඳුනාගන්න පුළුවන් columns තුනක් ලියන්න.

UNIT 09 / LESSON 04

Sampling bias - පෙනෙන දේම සියල්ල නෙවෙයි

අපේ CSV එකේ 2026 සැප්තැම්බර් දින 20ක් පමණයි. වෙනත් මාසවල උත්සව, පාසල් නිවාඩු හෝ වෙනත් වෙනස්කම් තිබිය හැක. ඒ නිසා 20 දිනක sample එකෙන් වසරක තත්ත්වය කියන්න එපා.

WATCH THE STEPS

1. Check dates
2. Spot missing context
3. Qualify



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / notes/bias.py

EXPLORER
data-analysis/
run_all.py
data/
shop_sales_raw.csv
lessons/
00_hello.py
outputs/
charts/
tests/

```
start = "2026-09-01"
end = "2026-09-20"
print("Sample window:", start, "to", end)
```

■ Windows Command Prompt / example output

```
Sample window: Sep 01 - Sep 20
```

REMEMBER

Sample එක සම්පූර්ණ ජනගහනය නියෝජනය නොකරන්නේ නම් conclusions සීමිතයි.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

මෙම කෙටි sample එකෙන් වසරේ “best month” එක තෝරාගන්න බැරි ඇයි?

UNIT 09 / LESSON 05

Correlation සහ causation වෙනස

දවස් දෙකක rainfall සහ juice sales එකට ඉහළ ගියා කියලා වැස්ස නිසාම juice අලෙවි වූ බව කිව නොහැක. එකට වෙනස්වීම සහ නිශ්චිත හේතුව එකම දේ නොවේ.

WATCH THE STEPS

1. Notice pattern
2. Find alternatives
3. Test



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / notes/causation.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
print("Correlation does not prove causation")
```

■ Windows Command Prompt / example output

```
Two things move together?
That alone is not proof of cause.
```

REMEMBER

Graph එකක් හේතුවක් ඔප්පු කරන්නේ නැහැ.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Sales වැඩි වුණු දවසකට හේතු දෙකක් අනුමාන කර ඒවා verify කරන්න දත්ත මොනවාද කියන්න.

UNIT 09 / LESSON 06

2026 දී Data Analyst skill map

Python basics, data cleaning, pandas, charts, SQL basics, clear writing සහ ethics යන skills එකට ඉගෙනගත් විට ඔබට data problems විසඳීමට හොඳ පදනමක් ලැබෙනවා. එක් tool එකක් පමණක් දැනීම ප්‍රමාණවත් නොවේ.

WATCH THE STEPS

1. Foundation
2. Practice
3. Explain



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / notes/roadmap.py

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
skills = ["Python", "pandas", "Charts",
          "SQL basics", "Communication", "Ethics"]
print(", ".join(skills))
```

■ Windows Command Prompt / example output

```
Python -> Data -> Charts -> Communication
```

REMEMBER

AI භාවිතයත් සමඟ accurate data checking සහ context පැහැදිලි කිරීම ඉතා වැදගත්.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

ඊළඟට ඉගෙනගන්න කැමති skill එකක් තෝරා හේතුව කියන්න.

UNIT 09 / LESSON 07

මෙම්ම mini project එක

School canteen එකක synthetic data හාවිතා කර product, quantity, price columns සකස් කරන්න. Personal names, phone හෝ private info නොදමන්න. CSV එක pandas මඟින් කියවා groupby කර bar chart එකක් අඳින්න.

WATCH THE STEPS

1. Make sample
2. Analyze
3. Present



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / project/challenge.csv

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
# Create fictional canteen data:
# item, quantity, price
# Banana, 2, 60
# Bun, 3, 90
# Juice, 1, 120
```

■ Windows Command Prompt / example output

```
Goal: canteen_report.csv
Goal: canteen_chart.png
```

REMEMBER

CSV data සහ chart දෙකම සමඟ conclusions සරලව explain කරන්න.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

Data rows 10ක් සාදා වැඩිම revenue product එක සොයාගන්න.

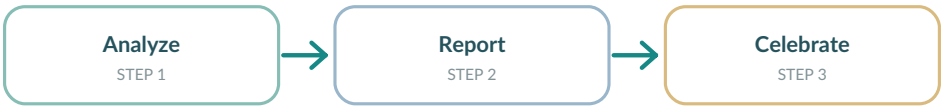
UNIT 09 / LESSON 08

අවසන් ඉගෙනුම් පරීක්ෂාව

Python run කළ හැකිද? CSV කියවිය හැකිද? missing values හඳුනාගත හැකිද? revenue sum සහ category groupby කළ හැකිද? Chart එක නිවැරදි units සමඟ explain කළ හැකිද? මේවා ඔබගේ final skills.

WATCH THE STEPS

1	2	3
Run project	Explain chart	Present



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

VS CODE / CODE EXCERPT

◆ Visual Studio Code / **CMD • final challenge**

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
python run_all.py
python -m unittest discover -s tests -v
```

■ Windows Command Prompt / example output

```
Done! Open charts/ and outputs/.
Ran 5 tests
OK
```

REMEMBER

වැරැද්දක් අහු වුණොත් ඒකත් ඉගෙනීමේ කොටසක්.

YOUR MINI MISSION / පුංචි අභ්‍යාසය

ඔබම තවත් සරල metric එකක් එකතු කර report එක update කරන්න.

UNIT 09 • THINK AND CHOOSE

කුඩා දැනුම් පරීක්ෂාව

වැරදුණොත් පසුගිය පාඩම නැවත කියවන්න. Answers අවසානයේ තිබෙනවා; මුලින්ම ඔබම උත්සාහ කරන්න.

1. AI එකක් totals ලිව්වොත්?

A. Blindly trust
B. Verify with data
C. Delete source

2. Private customer details?

A. Avoid sharing
B. Always post online
C. Include in tutorial

3. 20 days වලින් වසරම predict?

A. Reliable
B. Not justified
C. Always possible

කඩදාසියේ පිළිතුරු දෙන්න

1. මෙම Unit එකේ ඉගෙනගත් වැදගත්ම අදහස කුමක්ද?
2. මේ සතියේ ඔබම run කළ Python code එක කුමක්ද?
3. ඊළඟට වැඩිදුර පැහැදිලි කරගන්න අවශ්‍ය ප්‍රශ්නය කුමක්ද?

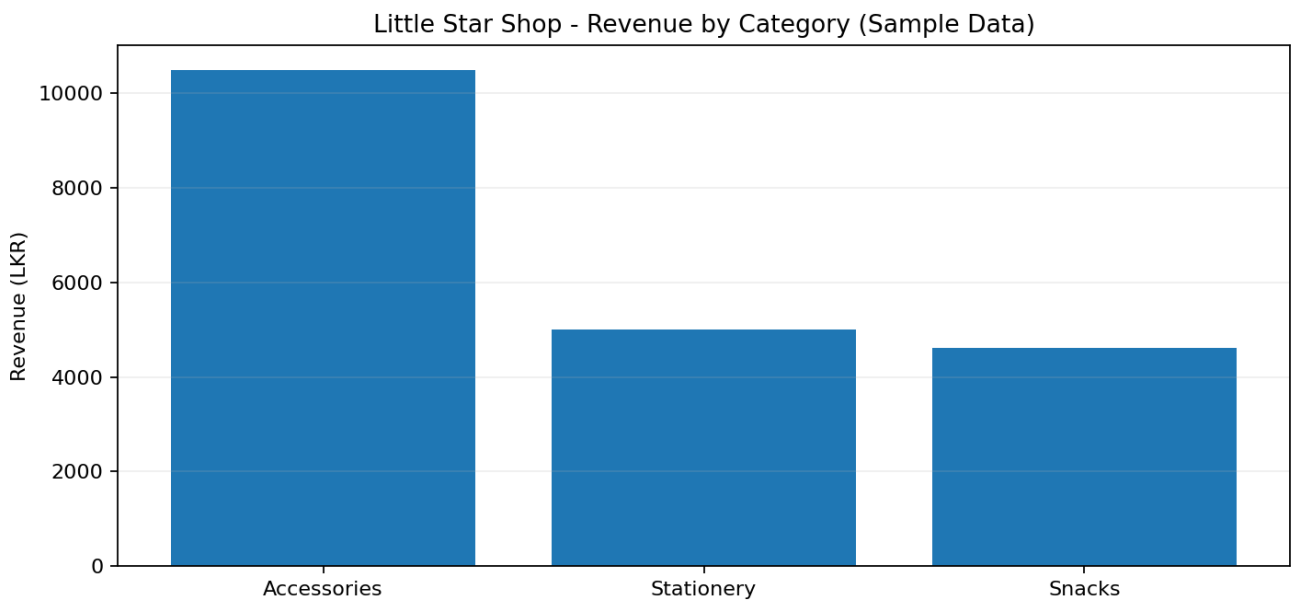
MINI MISSION

මිතුරෙකුට මිනිත්තු දෙකකින් අද පාඩමේ අදහස පැහැදිලි කරන්න.

FULL-PAGE GRAPH GALLERY

විකුණුම් කාණ්ඩ තුන සසඳමු

Category bar chart • example data



INTERPRETATION

Accessories revenue වැඩි වුණත් Profit ගැන තවම දත්ත නැහැ. Bars තුනේ units එක LKR බව බලන්න.

ගුරුවරයා සමඟ සාකච්ඡා කරන්න

Chart title කියවන්න. X-axis සහ Y-axis වල අර්ථ කියන්න. එක් නිවැරදි observation එකක් සහ එක limitation එකක් ලියන්න.

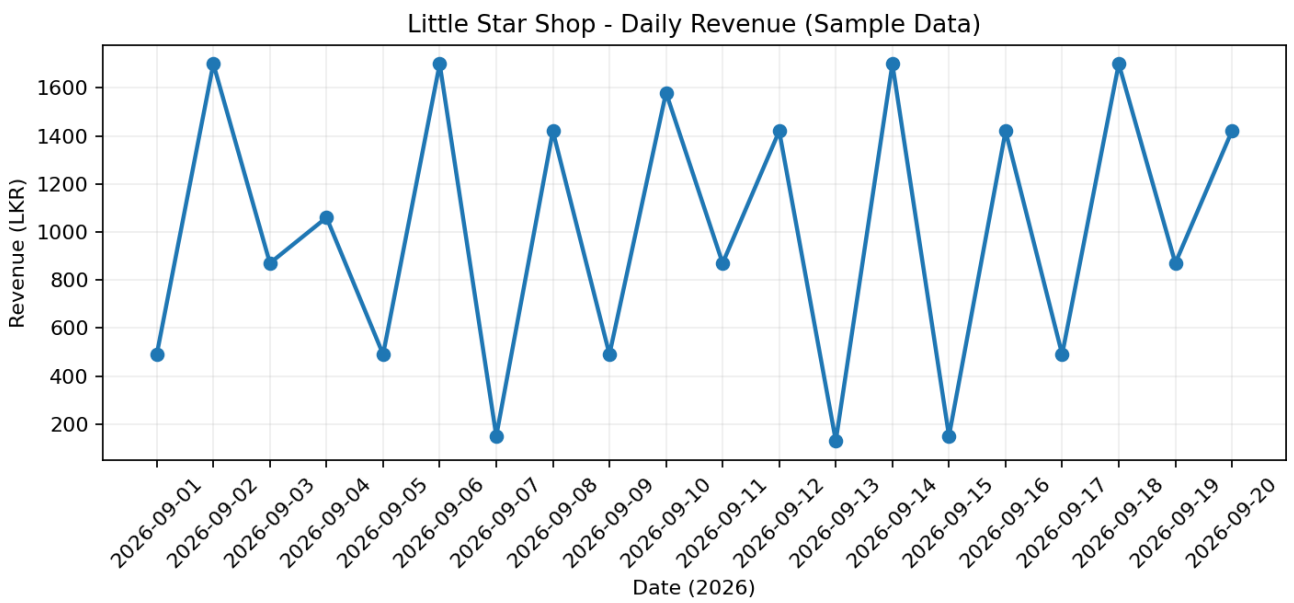
GRAPH MISSION

එකම graph එකේ labels මකා දැම්මොත් ඇතිවන ගැටලුව ලියන්න.

FULL-PAGE GRAPH GALLERY

දවසින් දවස වෙනස බලමු

Daily revenue line chart • example data



INTERPRETATION

මෙම line chart එකෙන් past data පමණක් පෙන්වයි. ඉදිරියේ විකුණුම් නියතව predict නොවේ.

ගුරුවරයා සමඟ සාකච්ඡා කරන්න

Chart title කියවන්න. X-axis සහ Y-axis වල අර්ථ කියන්න. එක් නිවැරදි observation එකක් සහ එක limitation එකක් ලියන්න.

GRAPH MISSION

එකම graph එකේ labels මකා දැම්මොත් ඇතිවන ගැටලුව ලියන්න.

7 DAY REAL PRACTICE PLAN

ඔබගේ පුංචි Data Analyst පුහුණුව

දවසට මිනිත්තු 20–40ක් වගේ කෙටි පුහුණුවක්. වැඩිහිටියෙකුගේ උදව් සමඟ අවශ්‍ය තරම් අවස්ථා ගන්න.

DAY 01

Python install කරන්න; 00_hello.py සහ variables.py run කරන්න.

DAY 02

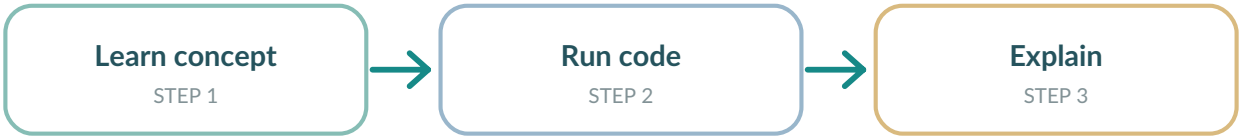
Lists, loops, functions ඉගෙනගෙන calculate_revenue() run කරන්න.

DAY 03

CSV file එකේ rows හඳුනාගන්න; 01_first_pandas.py run කරන්න.

DAY 04

Missing, duplicate, dates, quantity rules අධ්‍යයනය කරන්න.



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

◆ Visual Studio Code / CMD • final check

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
python run_all.py
python -m unittest discover -s tests -v
```

IMPORTANT

Quickly finishing pages is not the goal. Each command and chart meaning can be explained in your own words.

7 DAY REAL PRACTICE PLAN

ඔබගේ පුංචි Data Analyst පුහුණුව

දවසට මිනිත්තු 20–40ක් වගේ කෙටි පුහුණුවක්. වැඩිහිටියෙකුගේ උදව් සමඟ අවශ්‍ය තරම් අවස්ථා ගන්න.

DAY 05

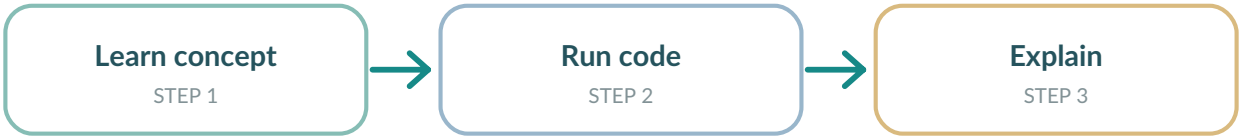
02_clean_sales.py සහ 03_analyze_sales.py run කර totals බලන්න.

DAY 06

04_make_charts.py run කර chart දෙක ගැන වාක්‍ය දෙකක් ලියන්න.

DAY 07

05_final_report.py සහ tests run කර mini presentation එකක් දෙන්න.



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

◆ Visual Studio Code / CMD • final check

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
python run_all.py
python -m unittest discover -s tests -v
```

IMPORTANT

Quickly finishing pages is not the goal. Each command and chart meaning can be explained in your own words.

CHECK YOUR OWN WORK

MCQ පිළිතුරු

මුලින්ම ඔබම පිළිතුරු තෝරාගත් පසු මෙය බලන්න. වැරදුණු ප්‍රශ්න සඳහා Unit එකේ අදාල පාඩම නැවත විමසන්න.

UNIT 01
1:B, 2:B, 3:A
දත්ත කියවීමේ රහස්

UNIT 02
1:B, 2:A, 3:B
Python මුල සිට

UNIT 03
1:A, 2:B, 3:B
CSV සහ project setup

UNIT 04
1:A, 2:A, 3:B
DataFrame හඳුනාගනිමු

UNIT 05
1:A, 2:B, 3:A
වැරදි දත්ත සකස් කරමු

මතක තබාගත යුතු දේ

Data cleaning කළ විට කොපමණ rows ඉවත් වුණාද? Category totals හරිද? Chart එකේ units මොනවාද? Revenue සහ Profit වෙන් කළාද? මේ ප්‍රශ්න අහන්න ඉගෙනගන්න.

■ Windows Command Prompt / example output

```
python -m unittest discover -s tests -v
Ran 5 tests
OK
```

CHECK YOUR OWN WORK

MCQ පිළිතුරු

මුලින්ම ඔබම පිළිතුරු තෝරාගත් පසු මෙය බලන්න. වැරදුණු ප්‍රශ්න සඳහා Unit එකේ අදාල පාඩම නැවත විමසන්න.

UNIT 06
1:A, 2:A, 3:B
ගණන් සහ groupby

UNIT 07
1:A, 2:A, 3:A
Charts හදා කියවමු

UNIT 08
1:A, 2:B, 3:A
Reports සහ checks

UNIT 09
1:B, 2:A, 3:B
AI කාලයේ Data Thinking

මතක තබාගත යුතු දේ

Data cleaning කළ විට කොපමණ rows ඉවත් වුණාද? Category totals හරිද? Chart එකේ units මොනවාද? Revenue සහ Profit වෙන් කළාද? මේ ප්‍රශ්න අහන්න ඉගෙනගන්න.

■ Windows Command Prompt / example output

```
python -m unittest discover -s tests -v
Ran 5 tests
OK
```

GET STARTED IN FIVE COMMANDS

ZIP එක Download කර Run කරන්න

ඔබගේ Windows computer එකේ Python install කර තිබේ නම් ZIP එක extract කර project folder එක VS Code එකේ open කරන්න. පසුව Terminal එකේ commands එකින් එක run කරන්න.

1. ZIP file එක extract කරන්න
2. project folder එක VS Code හි open කරන්න
3. CMD / Terminal එකෙන් commands ධාවනය කරන්න

◆ Visual Studio Code / CMD • quick start

```
EXPLORER
data-analysis/
  run_all.py
  data/

shop_sales_raw.csv
  lessons/
    00_hello.py
  outputs/
  charts/
  tests/
```

```
py -m venv .venv
.venv\Scripts\activate
python -m pip install -r requirements.txt
python run_all.py
python -m unittest discover -s tests -v
```

■ Windows Command Prompt / example output

```
Revenue (not profit): Rs 20,120.00
Units sold: 135
Ran 5 tests
OK
```

Output files මොනවාද?

data/ තුළ raw CSV එක; **outputs/** තුළ clean CSV, rejected rows, category/product/daily reports සහ summary; **charts/** තුළ PNG graphs දෙක.

> **PLATFORM NOTES**
>
> Mac/Linux commands, PowerShell විස්තර සහ troubleshooting README.md හි ඇත.

SOURCE CODE APPENDIX

data/create_sample_data.py

data/create_sample_data.py හි සම්පූර්ණ code එකේ කොටස 1/3. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / **data/create_sample_data.py**

```
"""Generate an entirely fictional shop dataset for the kids' tutorial."""
from pathlib import Path
import csv
from datetime import date, timedelta

BASE = Path(__file__).resolve().parent
OUTPUT = BASE / 'shop_sales_raw.csv'
PRODUCTS = [
    ('Pencil', 'Stationery', 40),
    ('Notebook', 'Stationery', 180),
    ('Eraser', 'Stationery', 30),
    ('Water Bottle', 'Accessories', 350),
    ('Juice', 'Snacks', 120),
    ('Biscuit', 'Snacks', 90),
]

def create_sample_file() -> Path:
    rows = []
    first_day = date(2026, 9, 1)
    for index in range(60):
        product, category, price = PRODUCTS[index % len(PRODUCTS)]
        # Every row is one sale line; these are not real customers.
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

SOURCE CODE APPENDIX

data/create_sample_data.py

data/create_sample_data.py හි සම්පූර්ණ code එකේ කොටස 2/3. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / **data/create_sample_data.py**

```
        rows.append({
            'order_id': f'S{index+1:03d}',
            'date': (first_day + timedelta(days=index // 3)).isoformat(),
            'product': product,
            'category': category,
            'quantity': str(1 + (index % 4)),
            'unit_price': str(price),
        })

    # Intentional beginner-friendly data problems to clean later.
    rows[4]['product'] = '  JUICE  '
    rows[8]['category'] = ' stationery '
    rows[11]['quantity'] = ''
    rows[19]['quantity'] = 'two'
    rows[28]['date'] = '2026-02-30'
    rows[37]['unit_price'] = '0'
    rows[43]['quantity'] = '-1'
    rows.append(rows[9].copy())  # duplicated order_id

    with OUTPUT.open('w', encoding='utf-8', newline='') as handle:
        writer = csv.DictWriter(handle, fieldnames=list(rows[0]))
        writer.writeheader()
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

SOURCE CODE APPENDIX

data/create_sample_data.py

data/create_sample_data.py හි සම්පූර්ණ code එකේ කොටස 3/3. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / **data/create_sample_data.py**

```
45          writer.writerows(rows)
46      return OUTPUT
47
48  if __name__ == '__main__':
49      print('Created:', create_sample_file())
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

> **RUN THIS FILE**
>
> Project root එකේ CMD මඟින් `python data/create_sample_data.py` run කර බලන්න. Dependencies නිවැරදිව install කර තිබිය යුතුයි.

SOURCE CODE APPENDIX

01_first_pandas.py

01_first_pandas.py හි සම්පූර්ණ code එකේ කොටස 1/1. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / **01_first_pandas.py**

```
"""Open the fictional sales CSV in a pandas DataFrame."""
from pathlib import Path
import pandas as pd

ROOT = Path(__file__).resolve().parent
sales = pd.read_csv(ROOT / 'data' / 'shop_sales_raw.csv')
print('First five rows:')
print(sales.head())
print('\nShape (rows, columns):', sales.shape)
print('\nMissing values by column:')
print(sales.isna().sum())
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

> **RUN THIS FILE**
>
> Project root එකේ CMD මඟින් `python 01_first_pandas.py` run කර බලන්න. Dependencies නිවැරදිව install කර තිබිය යුතුයි.

SOURCE CODE APPENDIX

02_clean_sales.py

02_clean_sales.py හි සම්පූර්ණ code එකේ කොටස 1/3. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / **02_clean_sales.py**

```
"""Clean fictional shop sales conservatively, without inventing missing values."""
from pathlib import Path
import pandas as pd

ROOT = Path(__file__).resolve().parent
SOURCE = ROOT / 'data' / 'shop_sales_raw.csv'
CLEAN = ROOT / 'outputs' / 'shop_sales_clean.csv'
REJECTED = ROOT / 'outputs' / 'rejected_rows.csv'


def clean_sales(data: pd.DataFrame) -> tuple[pd.DataFrame, pd.DataFrame]:
    """Return (valid sale lines, rejected rows) with a repeatable cleanup rule."""
    df = data.copy()
    df['product'] = df['product'].astype('string').str.strip().str.title()
    df['category'] = df['category'].astype('string').str.strip().str.title()
    df['date'] = pd.to_datetime(df['date'], errors='coerce', format='%Y-%m-%d')
    df['quantity'] = pd.to_numeric(df['quantity'], errors='coerce')
    df['unit_price'] = pd.to_numeric(df['unit_price'], errors='coerce')
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

SOURCE CODE APPENDIX

02_clean_sales.py

02_clean_sales.py හි සම්පූර්ණ code එකේ කොටස 2/3. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / 02_clean_sales.py

```
19
20      # Duplicate transaction identifiers could double-count sales.
21      duplicate = df['order_id'].duplicated(keep='first')
22      invalid = (
23          duplicate |
24          df[['date', 'product', 'category', 'quantity', 'unit_price']].isna().any(axis=1) |
25          (df['quantity'] <= 0) |
26          (df['unit_price'] <= 0) |
27          (df['quantity'] % 1 != 0)
28      )
29      rejected = df.loc[invalid].copy()
30      clean = df.loc[~invalid].copy()
31      clean['quantity'] = clean['quantity'].astype(int)
32      clean['revenue'] = clean['quantity'] * clean['unit_price']
33      clean['date'] = clean['date'].dt.strftime('%Y-%m-%d')
34      clean = clean.reset_index(drop=True)
35      return clean, rejected
36
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

SOURCE CODE APPENDIX

02_clean_sales.py

02_clean_sales.py හි සම්පූර්ණ code එකේ කොටස 3/3. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / 02_clean_sales.py

```
37
38  def main() -> None:
39      raw = pd.read_csv(SOURCE)
40      clean, rejected = clean_sales(raw)
41      CLEAN.parent.mkdir(exist_ok=True, parents=True)
42      clean.to_csv(CLEAN, index=False)
43      rejected.to_csv(REJECTED, index=False)
44      print(f'Original rows : {len(raw)}')
45      print(f'Clean rows    : {len(clean)}')
46      print(f'Rejected rows : {len(rejected)}')
47      print(f'Clean saved   : {CLEAN}')
48      print('Review rejected_rows.csv instead of silently guessing data.')
49
50
51  if __name__ == '__main__':
52      main()
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

> **RUN THIS FILE**
>
> Project root එකේ CMD මඟින් `python 02_clean_sales.py` run කර බලන්න. Dependencies නිවැරදිව install කර තිබිය යුතුයි.

SOURCE CODE APPENDIX

03_analyze_sales.py

03_analyze_sales.py හි සම්පූර්ණ code එකේ කොටස 1/3. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / **03_analyze_sales.py**

```
"""Group and summarize the cleaned fictional sales data."""
from pathlib import Path
import pandas as pd

ROOT = Path(__file__).resolve().parent
CLEAN = ROOT / 'outputs' / 'shop_sales_clean.csv'


def analyze_sales(df: pd.DataFrame) -> dict:
    revenue = float(df['revenue'].sum())
    category = (df.groupby('category', as_index=False)
                .agg(revenue=('revenue', 'sum'), units=('quantity', 'sum'))
                .sort_values('revenue', ascending=False))
    product = (df.groupby('product', as_index=False)
               .agg(revenue=('revenue', 'sum'), units=('quantity', 'sum'))
               .sort_values('revenue', ascending=False))
    daily = (df.groupby('date', as_index=False)
             .agg(revenue=('revenue', 'sum'))
             .sort_values('date'))
    return {
        'revenue': revenue,
        'lines': len(df),
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

SOURCE CODE APPENDIX

03_analyze_sales.py

03_analyze_sales.py හි සම්පූර්ණ code එකේ කොටස 2/3. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / 03_analyze_sales.py

```
        'units': int(df['quantity'].sum()),
        'average_line_revenue': float(df['revenue'].mean()),
        'category': category,
        'product': product,
        'daily': daily,
    }


def main() -> None:
    df = pd.read_csv(CLEAN)
    result = analyze_sales(df)
    for name in ('category', 'product', 'daily'):
        result[name].to_csv(ROOT / 'outputs' / f'{name}_report.csv', index=False)
    print(f"Revenue (not profit) : Rs {result['revenue']:,.2f}")
    print(f"Valid sale lines     : {result['lines']}")
    print(f"Units sold           : {result['units']}")
    print(f"Mean revenue / line  : Rs {result['average_line_revenue']:,.2f}")
    print('\nSales by category:\n', result['category'].to_string(index=False))
    print('\nTop product: ', result['product'].iloc[0]['product'])


if __name__ == '__main__':
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

SOURCE CODE APPENDIX

03_analyze_sales.py

03_analyze_sales.py හි සම්පූර්ණ code එකේ කොටස 3/3. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / 03_analyze_sales.py

```
45      main()
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

> **RUN THIS FILE**
>
> Project root එකේ CMD මඟින් `python 03_analyze_sales.py` run කර බලන්න. Dependencies නිවැරදිව install කර තිබිය යුතුයි.

SOURCE CODE APPENDIX

04_make_charts.py

04_make_charts.py හි සම්පූර්ණ code එකේ කොටස 1/2. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / **04_make_charts.py**

```
"""Create two accessible, labelled chart images for the shop report."""
from pathlib import Path
import matplotlib
matplotlib.use('Agg')  # Run without opening a GUI window.
import matplotlib.pyplot as plt
import pandas as pd

ROOT = Path(__file__).resolve().parent
CHARTS = ROOT / 'charts'
CHARTS.mkdir(exist_ok=True)


def make_charts() -> None:
    by_category = pd.read_csv(ROOT / 'outputs' / 'category_report.csv')
    by_day = pd.read_csv(ROOT / 'outputs' / 'daily_report.csv')

    # Chart 1: category comparison (revenue, not profit).
    fig, ax = plt.subplots(figsize=(9, 4.3), facecolor='white')
    ax.bar(by_category['category'], by_category['revenue'])
    ax.set_title('Little Star Shop - Revenue by Category (Sample Data)')
    ax.set_ylabel('Revenue (LKR)')
    ax.grid(axis='y', alpha=.2)
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

SOURCE CODE APPENDIX

04_make_charts.py

04_make_charts.py හි සම්පූර්ණ code එකේ කොටස 2/2. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / 04_make_charts.py

```
    fig.tight_layout()
    fig.savefig(CHARTS / '01_categories.png', dpi=160, facecolor='white')
    plt.close(fig)

    # Chart 2: daily trend, ordered by date.
    fig, ax = plt.subplots(figsize=(9, 4.3), facecolor='white')
    ax.plot(by_day['date'], by_day['revenue'], marker='o', linewidth=2)
    ax.set_title('Little Star Shop - Daily Revenue (Sample Data)')
    ax.set_ylabel('Revenue (LKR)')
    ax.set_xlabel('Date (2026)')
    ax.tick_params(axis='x', labelrotation=45)
    ax.grid(alpha=.2)
    fig.tight_layout()
    fig.savefig(CHARTS / '02_daily_trend.png', dpi=160, facecolor='white')
    plt.close(fig)
    print('Charts saved in:', CHARTS)


if __name__ == '__main__':
    make_charts()
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

> **RUN THIS FILE**
>
> Project root එකේ CMD මඟින් `python 04_make_charts.py` run කර බලන්න. Dependencies නිවැරදිව install කර තිබිය යුතුයි.

SOURCE CODE APPENDIX

05_final_report.py

05_final_report.py හි සම්පූර්ණ code එකේ කොටස 1/2. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / **05_final_report.py**

```
"""Write a plain text manager report using the current cleaned sample data."""
from pathlib import Path
import pandas as pd
import runpy

ROOT = Path(__file__).resolve().parent
# Numbers in filenames keep the tutorial steps in order.
analyze_sales = runpy.run_path(str(ROOT / '03_analyze_sales.py'))['analyze_sales']


def make_report() -> Path:
    sales = pd.read_csv(ROOT / 'outputs' / 'shop_sales_clean.csv')
    stats = analyze_sales(sales)
    leader = stats['product'].iloc[0]
    content = [
        '# Little Star Shop - Sample Sales Report (2026)',
      ",
        'All figures are invented for education. No customer data.',
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

SOURCE CODE APPENDIX

05_final_report.py

05_final_report.py හි සම්පූර්ණ code එකේ කොටස 2/2. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / **05_final_report.py**

```
        f"Valid sale lines: {stats['lines']}",
        f"Units sold: {stats['units']}",
        f"Total revenue: Rs {stats['revenue']:,.2f} (NOT profit)",
        f"Mean revenue per sale line: Rs {stats['average_line_revenue']:,.2f}",
        f"Highest revenue product: {leader['product']} (Rs {leader['revenue']:,.2f})",
      ",
        'Averages describe these sample rows only.',
        'Before making a business decision, check the source data.',
        'Rejected rows are saved for a human to review.',
    ]
    output = ROOT / 'outputs' / 'shop_report.txt'
    output.write_text('\n'.join(content) + '\n', encoding='utf-8')
    return output


if __name__ == '__main__':
    print('Report saved:', make_report())
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

> **RUN THIS FILE**
>
> Project root එකේ CMD මඟින් `python 05_final_report.py` run කර බලන්න. Dependencies නිවැරදිව install කර තිබිය යුතුයි.

SOURCE CODE APPENDIX

run_all.py

run_all.py හි සම්පූර්ණ code එකේ කොටස 1/1. ZIP එකේ file එක පූර්ණයෙන්ම තියෙනවා; PDF එකේ line numbers කියවීම සඳහා පමණයි.

◆ Visual Studio Code / **run_all.py**

```
1  """Create data, clean it, analyze it, chart it, and write the report."""
2  from pathlib import Path
3  import subprocess
4  import sys
5
6  ROOT = Path(__file__).resolve().parent
7  FILES = [
8      'data/create_sample_data.py',
9      '01_first_pandas.py',
10     '02_clean_sales.py',
11     '03_analyze_sales.py',
12     '04_make_charts.py',
13     '05_final_report.py',
14 ]
15 for file_name in FILES:
16     print('\n===', file_name, '===', flush=True)
17     subprocess.run([sys.executable, str(ROOT / file_name)], check=True, cwd=ROOT)
18 print('\nDone! Open charts/ and outputs/.')
```

TIP: මෙම පිටුවේ line numbers code එකට copy කරන්න එපා. ZIP එකේ original script භාවිතා කරන්න. Python code comments English භාෂාවෙන්.

> **RUN THIS FILE**
>
> Project root එකේ CMD මඟින් `python run_all.py` run කර බලන්න. Dependencies නිවැරදිව install කර තිබිය යුතුයි.

WORDS TO REMEMBER

පුංචි Data Dictionary

ඉගෙනුමේදී අලුත් වචන අමතක වීම සාමාන්‍යයි. මේ sheet එක හැමවිටම නැවත බලන්න පුළුවන්.

Data
තොරතුරු/අගයන්

Row
එක් වාර්තාවක්

Column
එකම වර්ගයේ තොරතුරු තීරුවක්

CSV
Comma-separated-values text file

DataFrame
pandas වගු ව්‍යුහය

Missing value
නොදන්නා හෝ හිස් අගය

Duplicate
නැවත ලියා ඇති වාර්තාව

Revenue
විකුණුම් වලින් ලැබුණු මුළු මුදල

WORDS TO REMEMBER

පුංචි Data Dictionary

ඉගෙනුමේදී අලුත් වචන අමතක වීම සාමාන්‍යයි. මේ sheet එක හැමවිටම නැවත බලන්න පුළුවන්.

Profit
වියදම් ආදී දේ සලකා ලාභය

Mean
එකතුව / ගණන

Median
පිළිවෙළට තැබූ values වල මැද අගය

groupby
කණ්ඩායම් අනුව එකතු/සාරාංශ

Bar chart
categories සසඳන graph

Line chart
කාලයත් සමඟ වෙනස්වීම

Synthetic data
මනඃකල්පිත/අභ්‍යාස දත්ත

AI hallucination
AI විසින් අසත්‍ය තොරතුරු සත්‍ය වගේ කියන අවස්ථාව

OFFICIAL LEARNING REFERENCES • 2026

වැඩිදුර ඉගෙනීමට

මෙම පොතේ coding steps Python, pandas සහ Matplotlib හි සාමාන්‍ය API භාවිතා කරයි. Software updates නිසා versions සහ installer screens වෙනස් වෙන්න පුළුවන්. පහත නිල documentation බලන්න.

Python official tutorial
https://docs.python.org/3/tutorial/

Python packaging: pip + virtual environments
https://packaging.python.org/en/latest/guides/installing-using-pip-and-virtual-environments/

pandas getting-started tutorials
https://pandas.pydata.org/docs/getting_started/intro_tutorials/

Matplotlib quick start
https://matplotlib.org/stable/users/explain/quick_start.html

මෙම PDF ගැන

2026 සඳහා සකස් කළ සිංහල beginner educational guide එකකි. Charts සහ numbers sample project එකෙන් generated; වෙනත් සමාගමක sales data නොවේ. Product names, prices සහ sample dates අධ්‍යාපනික fictional values ය.

> **EGOTECHWORLD**
> Keep building small projects and explaining your findings with care.

GREAT WORK, DATA DETECTIVE!

මගේ පුංචි Data Analyst සටහන

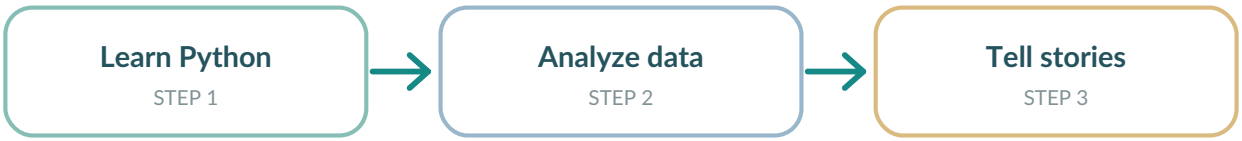
මෙම page එක print කර ඔබගේ පුහුණු වාර්තාව ලියාගන්න. මේක ස්වයං අධ්‍යයන worksheet එකක්; විධිමත් වෘත්තීය සහතිකයක් නොවේ.

මගේ nickname: _____

Project එක run කළ දිනය: _____

මම ඉගෙනගත් හොඳම දෙය: _____

ඊළඟ project අදහස: _____



සරල සටහන: පියවර 1 → පියවර 2 → පියවර 3

> Data අගයන් පමණක් නොවේ. ඒවා තේරුම්ගෙන සාක්ෂි සහිතව කතා කිරීමයි!