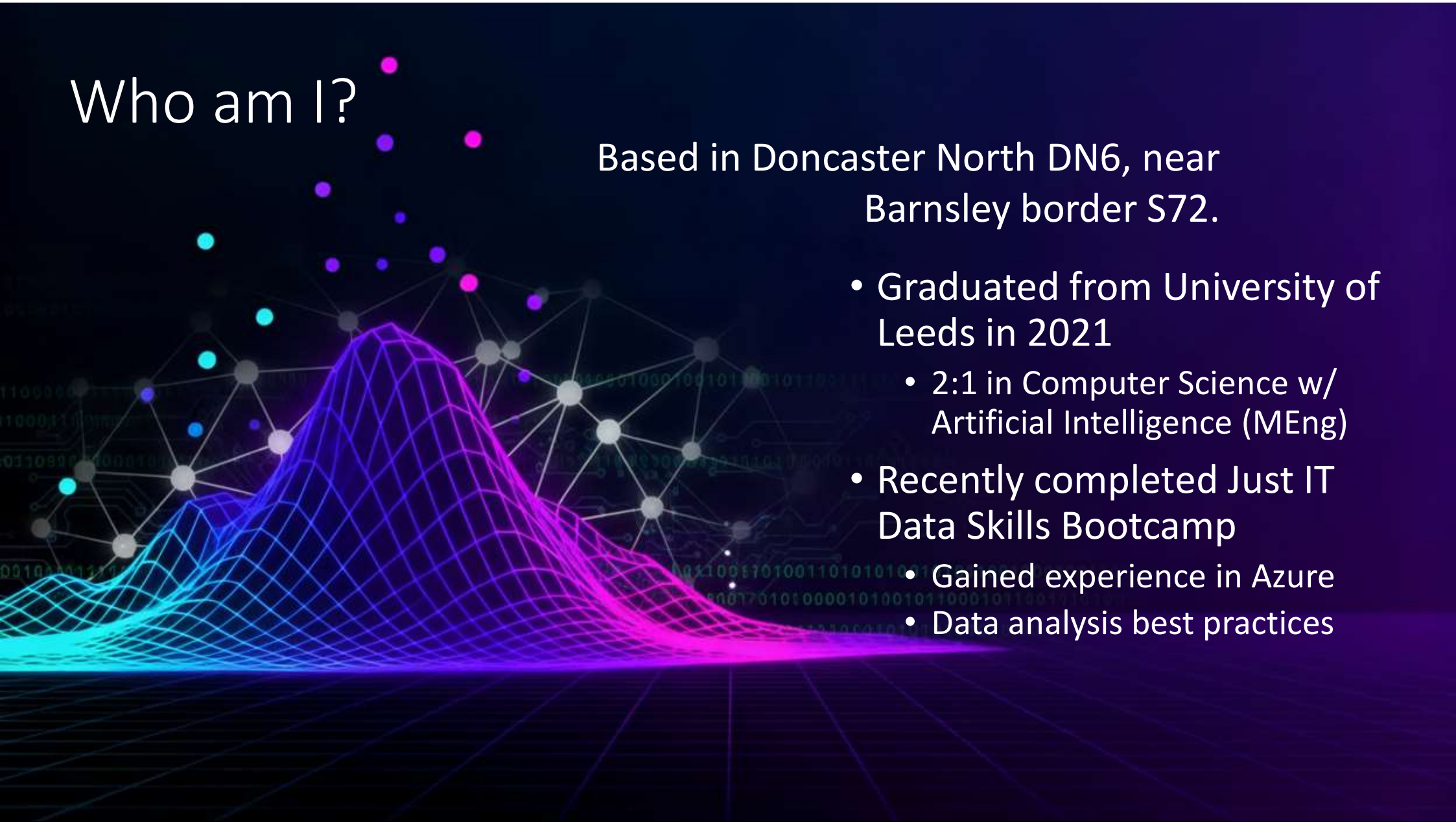
Jack Auty

Data Analyst Portfolio

Who am I?



Based in Doncaster North DN6, near Barnsley border S72.

- Graduated from University of Leeds in 2021
  - 2:1 in Computer Science w/ Artificial Intelligence (MEng)
- Recently completed Just IT Data Skills Bootcamp
  - Gained experience in Azure
  - Data analysis best practices

Background – Fun Facts

- I value my friends and relationships more than anything else
  - Strong sense of morality which leads me to attentiveness
  - An unshakeable need to defend and advocate for others

- Huge and life-long fan of the hit Manga, One Piece!

Background – Technical

- Always loved exploring technology
  - Tinkering with Operating Systems at a young age
  - Running Minecraft Servers

This curiosity lead me to exploring SQL, Cloud Computing, Data Science and Machine Learning at University

Background – Work History

You may have noticed the gap between my graduation and completion of this program…

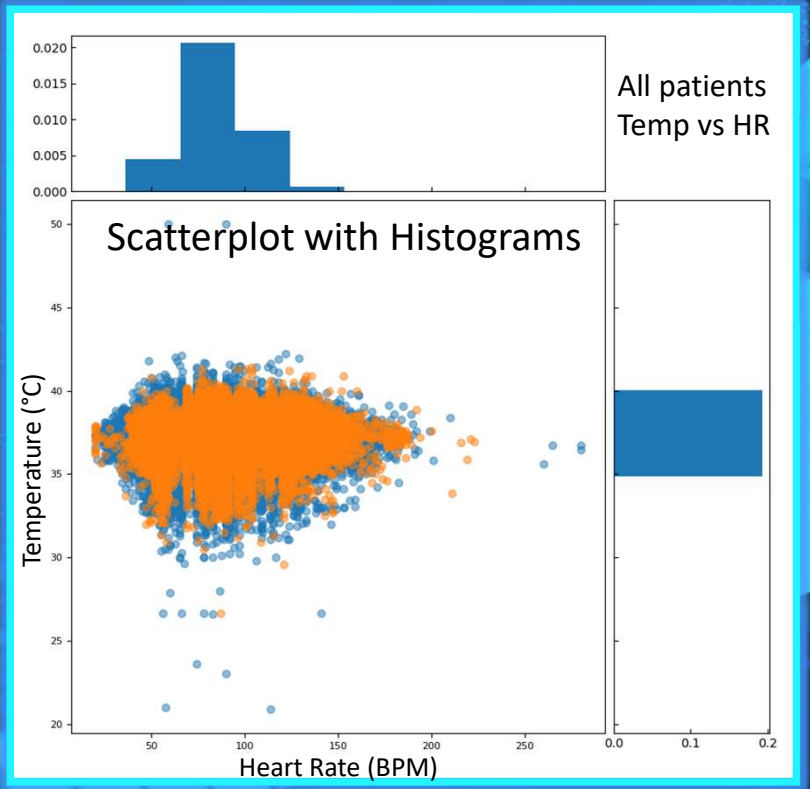
- Family commitments
- Completed Foundational C# O'Reilly Course
- Diagnosed and titrated onto stable medication for ADHD
- Few work experiences, primarily academic focus

Looking to secure a role in Data Analysis and these projects demonstrate the necessary skillset

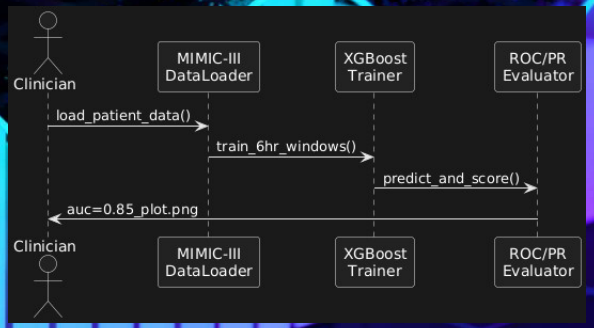
github.com/JackEdwardAuty/Bachelors-Project-Early-Prediction-of-Sepsis-from-Clinical-Data

Early Prediction of Sepsis from Clinical Data

- Bachelor's Project
- Important to detect sepsis early as vital to improving clinical outcomes.

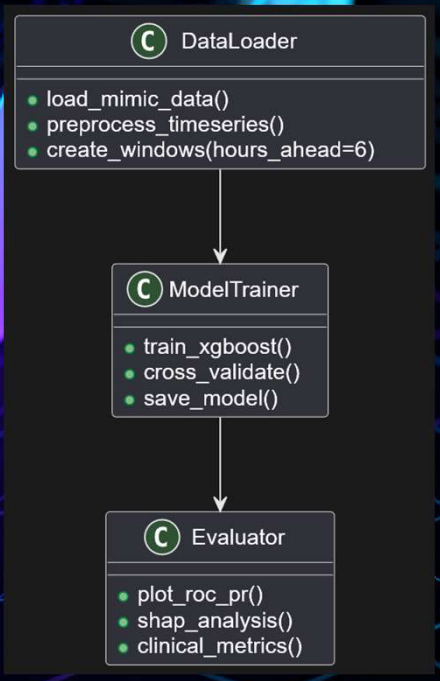


← Old Graph created with matplotlib



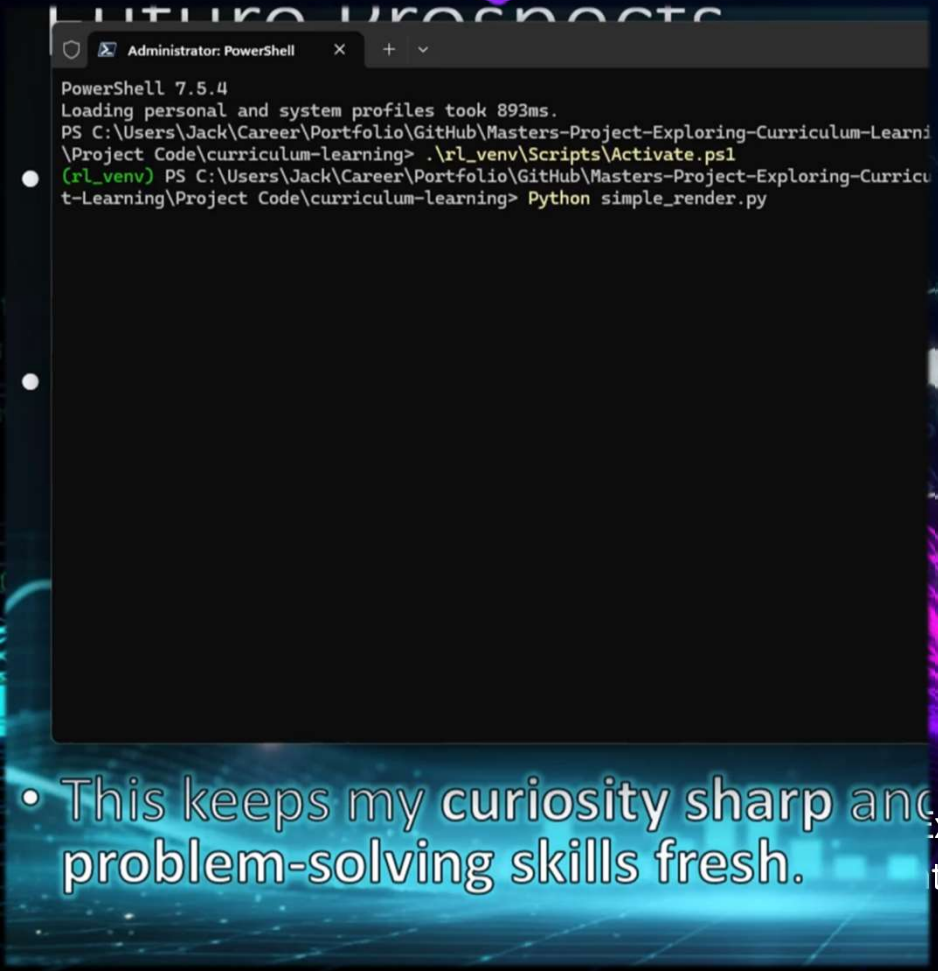
Self-Directed (+ Programmatic) Critical Evaluation revealed project issues but bolstered my knowledge

PyTorch implementation



Extracted trends from ICU Blood Test results that allowed earlier prediction than typical methods allow.

Exploring Curriculum Learning to improve Reinforcement Learning Outcomes



- This keeps my curiosity sharp and problem-solving skills fresh.

I'm sure we've all followed a curriculum before?

…Do AI agents learn better when taught in this way?

First stage of Curriculum:

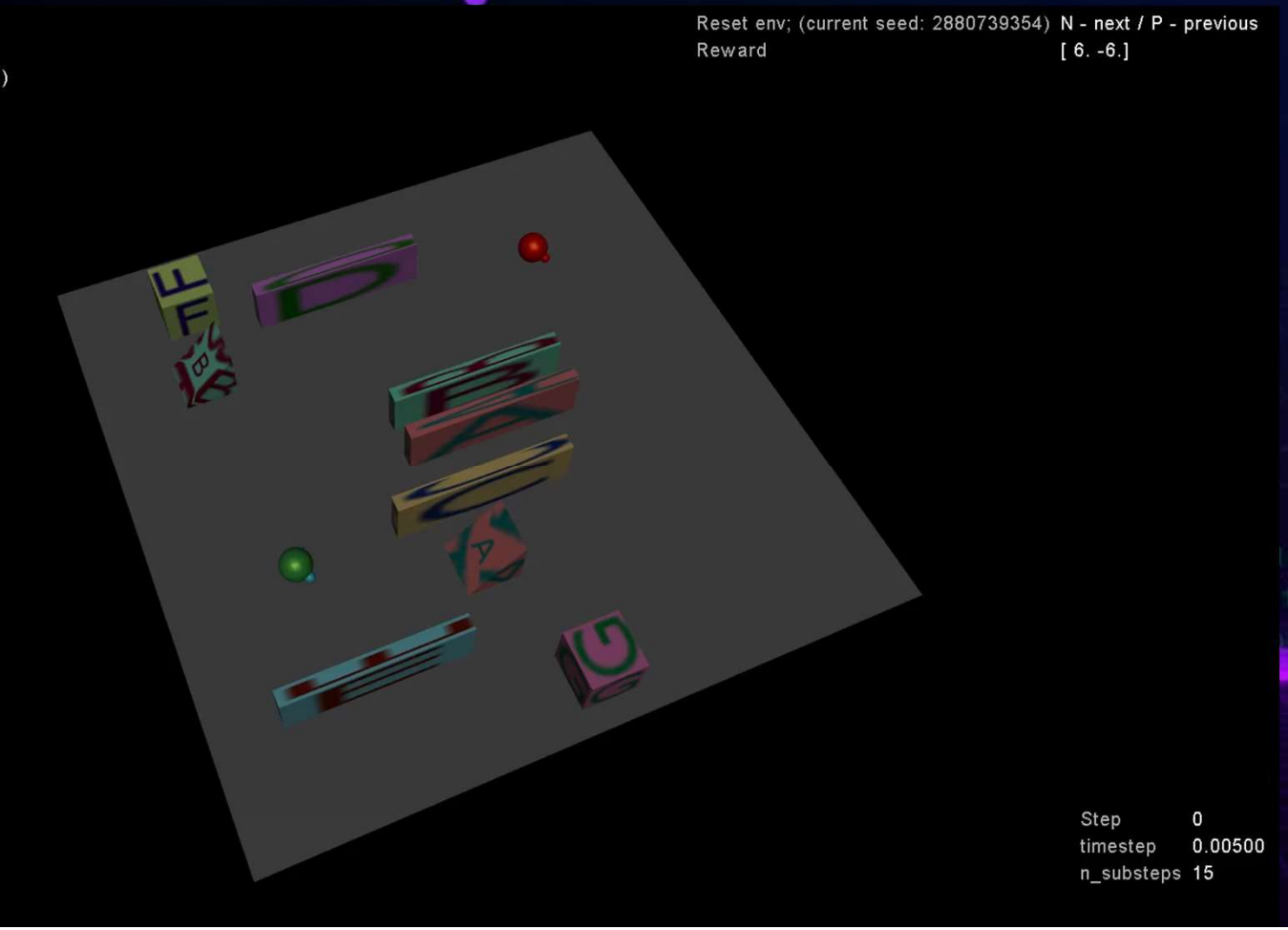
- Single Chaser and Target (live demo)

TensorFlow implementation with

NumPy for calculations

Matplotlib to render visualisations

Exploring Curriculum Learning to improve Reinforcement Learning Outcomes



…Do AI agents learn better when taught with a curriculum?

Complexity of RL environment is **exponential** with added obstacles and agents:

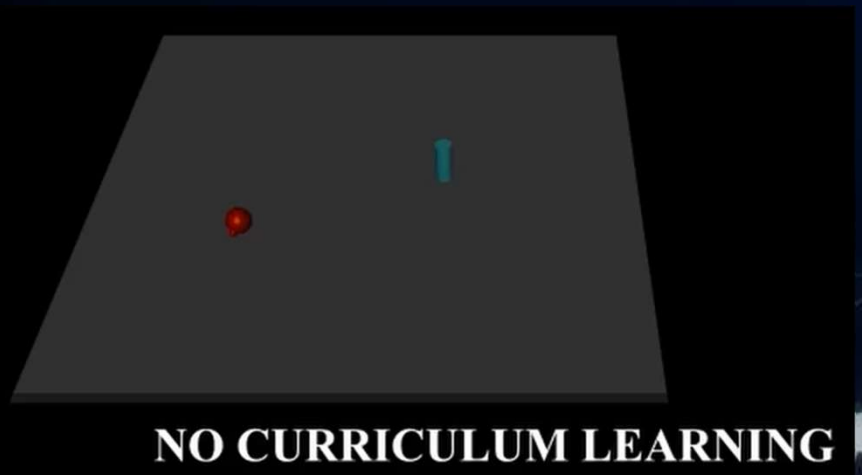
- Used to simulate and prepare for relief scenarios such as natural disasters, where the goal is clear but the path towards that goal can vary.

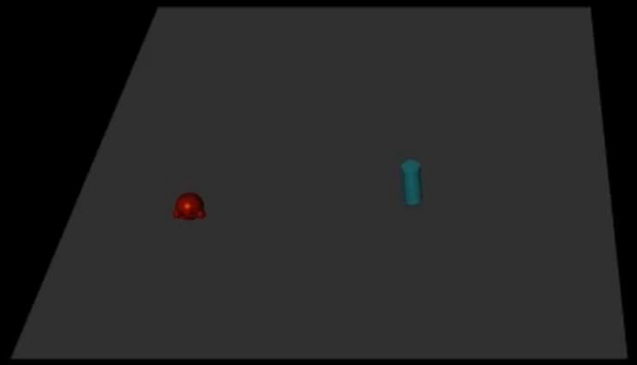
Introduced stationary and agent adjustable obstacles:

- Eventually developed into full Hide-and-Seek (video)

github.com/JackEdwardAuty/Masters-Project-Exploring-Curriculum-Learning-to-Improve-Reinforcement-Learning

Exploring Curriculum Learning to improve Reinforcement Learning Outcomes





github.com/JackEdwardAuty/Masters-Project-Exploring-Curriculum-Learning-to-Improve-Reinforcement-Learning

Increasing environment difficulty over time, rather than having agent tackle max complexity at all times.

Side-by-side demonstration of
Standard Reinforcement Learning
vs Curriculum Learning

…Do AI agents learn better when taught with a curriculum?

Answer… YES!

Reinforcement Learning agents also learned ‘better’ – more quickly and efficiently – *when following a Curriculum*.

Where did this leave me?

- Strong Analytical Skills developed
- Deep appreciation for the **power of data** – how it can be used to solve real-world problems

Graduated during Pandemic

Struggled with Confidence, Direction, Imposter Syndrome

Turning Point – Data Skills Bootcamp

Reignited my drive to enter the industry

- Inspiring quality of teaching
- Warmth, Resilience and Insight of cohort

Confident in handling Extract, Transform, and Load process along with Data Visualisation and Analysis!

Competency Demonstration:

Markdown Processing Python Personal Project

github.com/JackEdwardAuty/Markdown-Notes-Processing-and-Analysis/

Markdown Processing Personal Project

Intro – Data Collection & Tools

- Built in **Python**, utilising **DataFrame**, **Pandas** libraries
- **Markdown**: industry standard lightweight markup language used for documenting code bases, processes and readme files.
- **Obsidian Vault:** Custom markdown-based PKM system
  - templated daily entries for medication details, and wake hours
- Custom File-Handling and *Data **Extract**ion* Pipeline
  - Custom Regex to handle varying data entry formats across Templates

Personal-Knowledge-Management

```
48 extract_wake_times(process_file_headers(*get_file_info(test_case_full)),
     get_file_info(test_case_full))
49
```

```
[{'Sleep': '~',
  'Wake': '14:45',
  'Time': '',
  'Quality': 'Decent',
  'date': '2025-11-30',
  'filename': '2025-11-30.md',
  'folder': '2025-11',
  'section_type': 'Wake+Sleep'}]
```

Wake

Table of Contents

Sleep

Sleep	Wake	Time	Quality
~	{{time}}		Decent

```
Processing 1513 markdown files...
Saved wake_time.csv with 118 rows
Saved medication.csv with 901 rows
```

```
53 extract_med_times(process_file_headers(*get_file_info(test_case_full)),
     get_file_info(test_case_full))
```

```
[{'Time': '17:09',
  'Medication': 'Ritalin IR',
  'Dosage': '5mg',
  'date': '2025-11-30',
  'filename': '2025-11-30.md',
  'folder': '2025-11',
  'section_type': 'Prescription'},
 {'Time': '17:09',
  'Medication': 'Fexofenadine',
```

Medication

Table of Contents

Prescribed

Time	Medication	Dosage
{{time}}	Ritalin XR	2x18mg
	Ritalin IR	5mg

github.com/JackEdwardAuty/Markdown-Notes-Processing-and-Analysis/

Markdown Processing Personal Project

- Custom *Data* ***Extract****ion* & ***Normalization*** Pipeline
  - **Process:** Additional bespoke regex patterns for Normalisation parse_dosage_string
  - Pandas transformations in Python; output standardized format for Tableau.

- ***Transform****ing* the extracted data to ease further analysis
  - **Non-Standard Formats Handled:**
    - **Dosage:**
      - 2x18mg → 36mg, [Ritalin XR]
      - 1 (tab) → 500mg, 1.5x500mg → 750mg [Paracetamol]
    - **Multiple Times:**
      - 15:25, 19:40 → separate rows (same dosage)
    - **Partial Doses:**
      - 17:15, 21:35 (half) → separate rows (adjust dosage)
    - **Time Ranges:**
      - 09:45-50 → 09:47 (midpoint)
    - **Non-Standard:**
      - logged to notes column for manual review

To implement these Transformations, I introduce three new methods:

- parse_dosage_string(dosage_str, factor=1.0)
  - Seven Regex patterns to handle varied input form
- parse_time_value(time_str)
  - Handles time ranges and non-standard to notes column
- split_times_to_rows(dataframe)
  - Handles splitting across multiple rows with multiplicative factors

```
def parse_dosage(dosage_str, factor=1.0):
    """
    Handles:
     - 2x18mg, 1.5x5mg, 1.5x500mg
     - 5mg, 120mg, 500mg
     - 2 x tab, 2 tab, 2 tabs
     - 1 (tab), 2 (tabs)
    Returns (amount, unit) or (None, original) if no match.
    """
```

- CSV now files ready for **Load** stage
  - Tableau for Visualisation and further Analysis

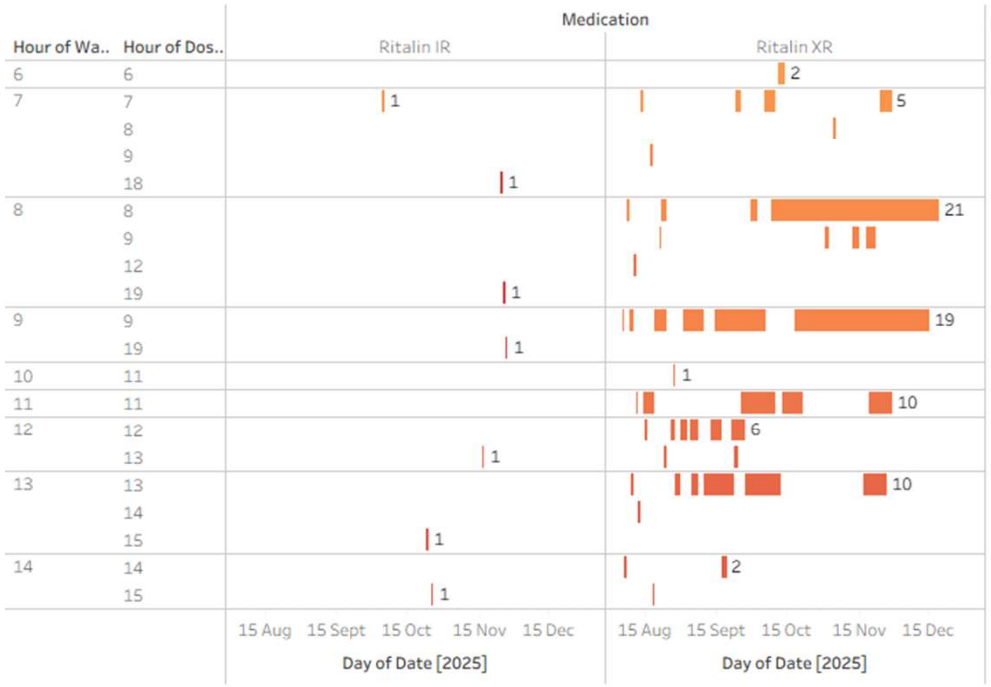
Markdown Processing Tableau Dashboard

Medication Timing, Wake Patterns, and Adherance Streaks

github.com/JackEdwardAuty/Markdown-Notes-Processing-and-Analysis/

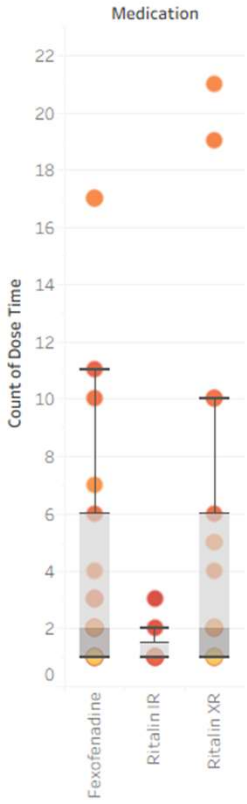
Exploring how my Wake Time and Dose Timing relate to Ritalin Adherance

Dose Time vs Wake Time (Prescription)



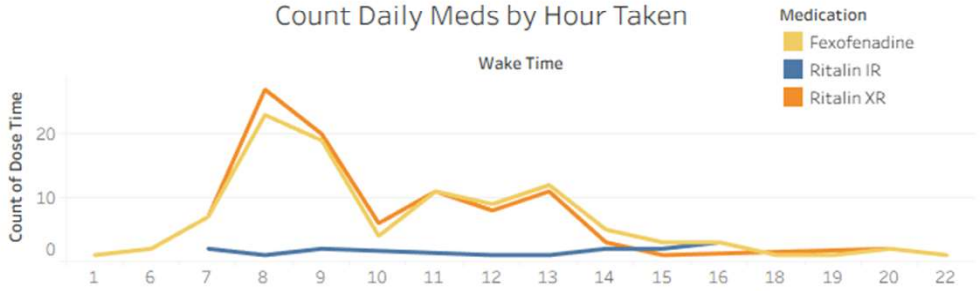
Dose Hour

Variation in Dose Hour

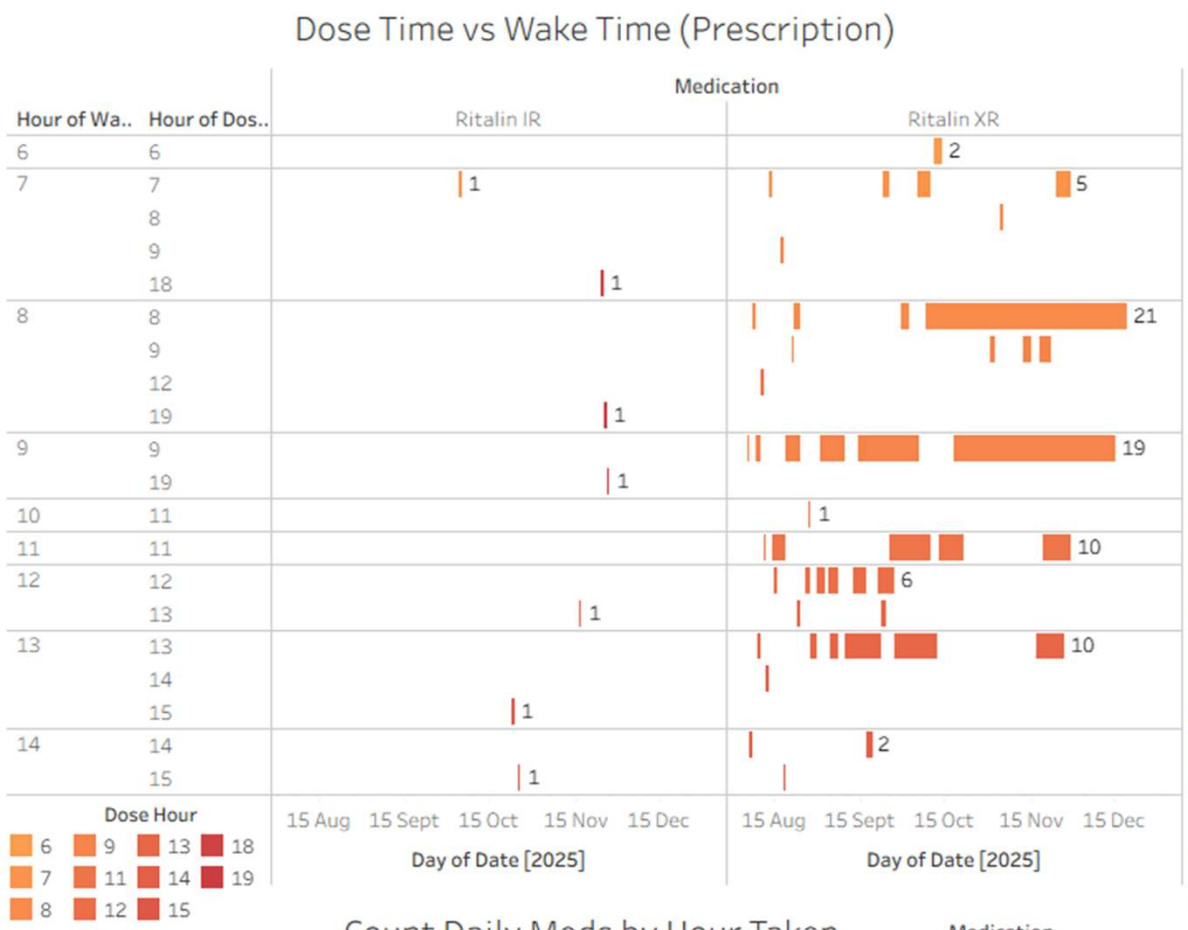


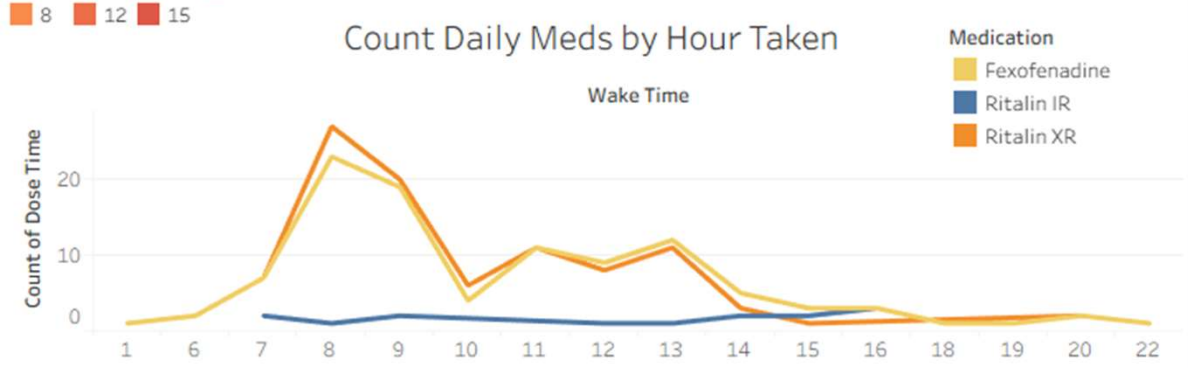
Streak Counter by Streak Band (Prescription)

Count Daily Meds by Hour Taken



Medication Streak by Wake Hour

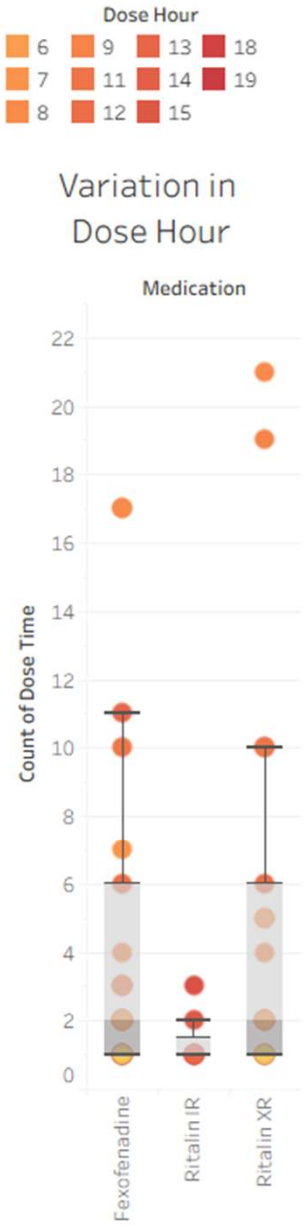


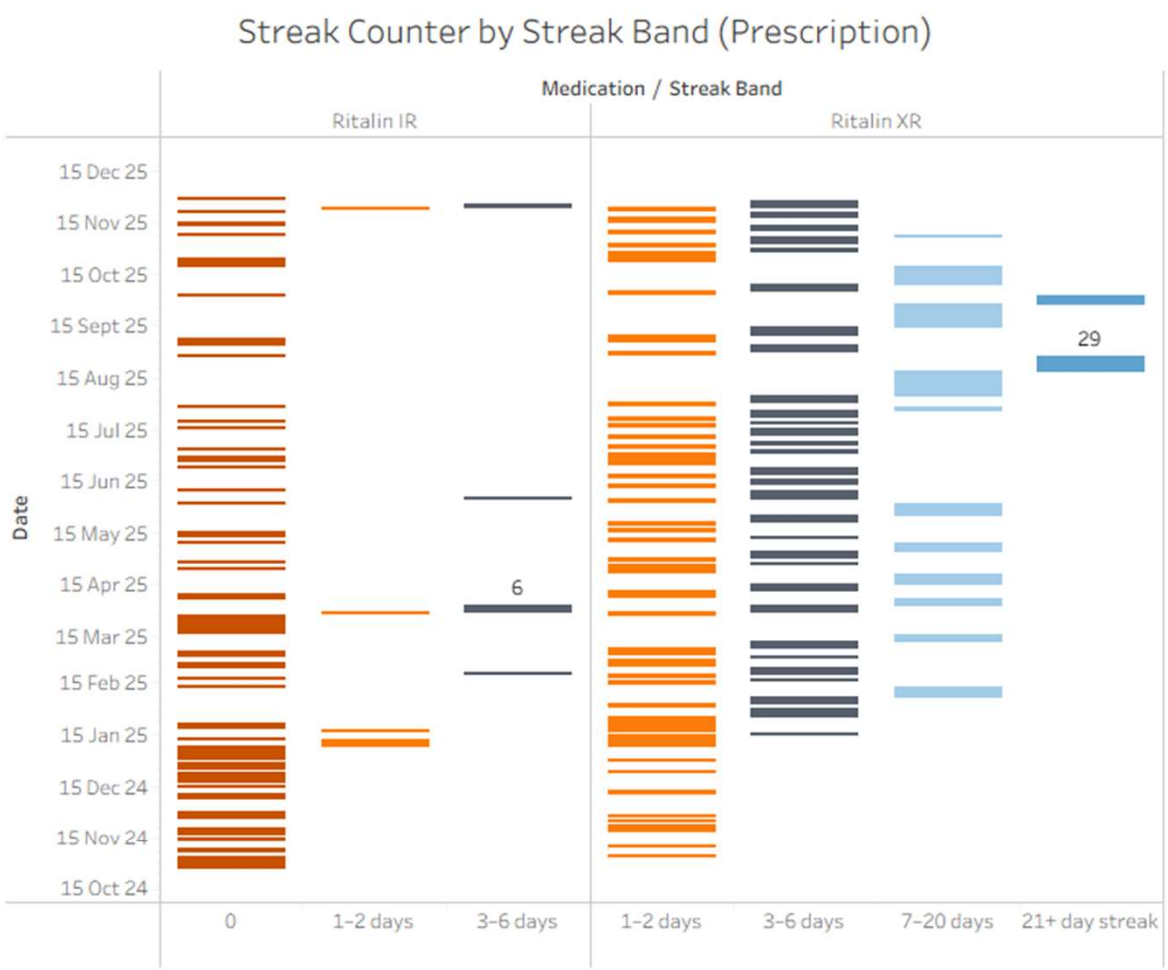


Dashboard Takeaways

- Dose Time against Wake Time Using RUNNING_SUM method
  - Found that most consistently took XR medication at 8 AM and 9 AM
  - More likely to take (late-day) Ritalin IR the earlier I woke up
- Count Daily Meds by Hour Taken
  - Stayed consistent with Fexofenadine whether woke up too late for Ritalin XR
  - Confirmation of 8-9 AM consistency

github.com/JackEdwardAuty/Markdown-Notes-Processing-and-Analysis/





Medication Streak by Wake Hour

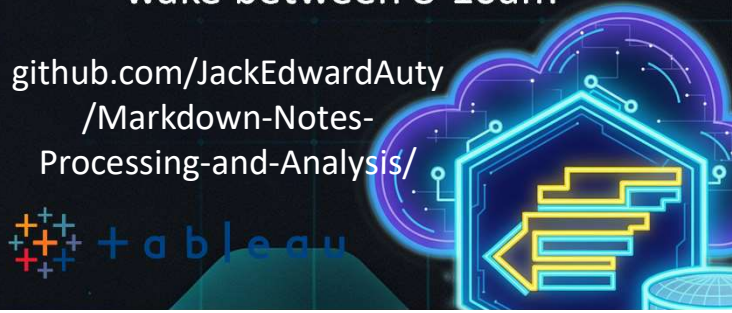
Medication	Streak Band	Wake Time 1	6	7	8	9	10	11	12	13	14	15	20	22
Ritalin XR	1-2 days			5	7	6	2		1	4		1		
	3-6 days			3	46	37	11	3	9	9			9	
	7-20 days		15	47	97	88	7	108	43	49	35			11
	21+ day streak	25			66	21	51	72	51	54				

Streak Band
0
1-2 days
3-6 days
7-20 days
21+ day streak

Dashboard Takeaways

- Variation in Dose Hour
  - Box-and-Whisker showing distribution
- Streak Counter by Band
  - Longest XR medication streak was 29 days in a row
  - Ritalin IR was rarely taken multiple days in a row
- Med Streak by Wake Hour
  - Longest runs (most consistent distribution) occurs when I wake between 8-10am

github.com/JackEdwardAuty/Markdown-Notes-Processing-and-Analysis/



Future Prospects

Connect with me:
GitHub: @jackedwardauty
Email: jackedwardauty@gmail.com

- Eager to begin career in Data Analysis or Business Intelligence
  - Looking to apply my analytical mindset, love of learning and technical expertise to deliver **meaningful** insights.
- Outside of work I'm a dedicated tinkerer
  - Open-Source Software
  - Operating System modification
  - PC Building (hardware)
  - Data-driven habit tracking
- This keeps my **curiosity sharp** and **problem-solving skills fresh.**