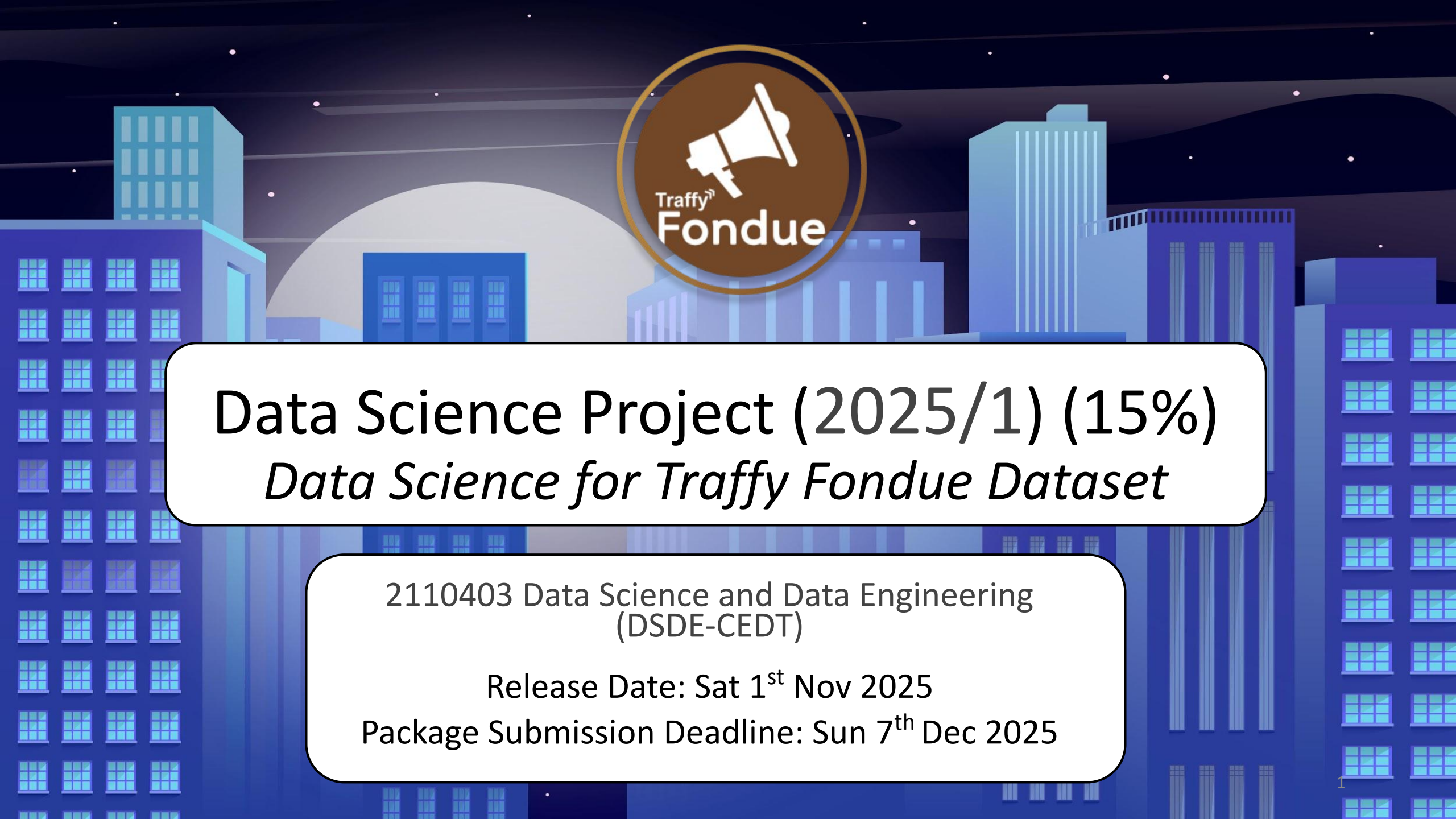

Data Science Project (2025/1) (15%)

Data Science for Traffy Fondue Dataset

2110403 Data Science and Data Engineering (DSDE-CEDT)

Release Date: Sat 1st Nov 2025

Package Submission Deadline: Sun 7th Dec 2025

Full raw Traffy Fondue

Data from Traffy Fondue Resources (Aug 2021 - Jan 2025),
Data consists of complaint reports submitted by citizens, primarily aggregated from Bangkok. The data are in CSV format
The data are in CSV format Data Explanation

- ticket_id: A unique identifier for each ticket or case.
- type: The category or type of issue reported in the ticket [multiple labels].
- organization: The organization or department handling the ticket.
- comment: complaints or feedback provided by users regarding public services.
- photo: A link to photos that support the issue.
- photo_after: A link to a photo taken after the issue has been addressed.
- coords: Coordinates (latitude and longitude).
- address: The physical address.
- subdistrict: The sub district in which the issue was reported.
- district: The district in which the issue was reported.
- province: The province in which the issue was reported.
- timestamp: The date and time when the ticket was created.
- state: The current state or status of the ticket.
- star: A numerical rating (0-5).
- count_reopen: The number of times the ticket was reopened.
- last_activity: The date and time of the last activity on the ticket.

```
Data columns (total 16 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   ticket_id      778254 non-null  object
 1   type           786929 non-null  object
 2   organization   786455 non-null  object
 3   comment        778254 non-null  object
 4   photo          786911 non-null  object
 5   photo_after    641309 non-null  object
 6   coords         787026 non-null  object
 7   address        778254 non-null  object
 8   subdistrict    786460 non-null  object
 9   district       786465 non-null  object
 10  province       786831 non-null  object
 11  timestamp      787026 non-null  object
 12  state          787026 non-null  object
 13  star           274097 non-null  float64
 14  count_reopen   787026 non-null  int64
 15  last_activity  787026 non-null  object
```

Data link:
https://drive.google.com/file/d/19QkF8i1my99gjbyHe7de_qZNwgrca6R5/view?usp=sharing


Tha Khlong
Khlong Luang
Thanyaburi
Lam Sai
Bang Nam Priao
Sam Khok
Lam Luk Ka
Lat Sawai
Lam Sam Kaeo
RANGSIT
Pathum Thani
Bang Phun
Lak Hok
Khu Khot
Nakhon Nueang Khet
Chachoengsao
Bang Luang
PAK KRET
Bang Bua Thong
NONTHABURI
Sao Thong Hin
Bang Yai
Bang Kruai
Plai Bang
Sala Ya
Bang Phli Noi
Hom Sin
Bang Bo
Bang Sao Thong
Bang Phli
Dan Samrong
Thepharak
Phra Pradaeng
Bang Mueang
SAMUT PRAKAN
Khlong Dan
Laem Fa Pha
Bang Pu
OM NOI
Bang Namchuet
Ban Nong Hat Yai
Ban Khok Kham
Ban San Dap
Ban Laem Yai
Ban Khlong Thale
Ban Sakhla
Ban Samut Chin
09/24/2021 – 01/16/2025
timestamp
Y Axis
October
2022
April
July
October
2023
April
July
October
2024
April
July
October
2025

Categories: คลอง,สัตว์จรจัด,ความสะอาด
Comment: ม.ศิรินครพัฒนา มวมินท์ 24 คลองกุ่ม บึงกุ่ม กทม หมาจรจัดช่วยมาจับด้วย และขยะมีคนมาทิ้งที่ว่างเปล่าพอมีทางช่วยไม่ครับ
Photo:

Date: 2022-06-02

Categories: ความสะอาด
Comment: ขอให้ทางเขตย้ายถังขยะด้วยค่ะ
Photo:

Date: 2022-06-01

Categories: ถนน,เสียงรบกวน,ความสะอาด
Comment: ปัญหารบกวนจากอาคาร BJC ซอยบาโบส2 สุขุมวิท42: 1) เสียงรบกวน- 1.1 การทำงานของรปภดีก มีพฤติกรรมใช้เสียงดังรบกวนอาคารรอบข้างทั้งเสียงตะโกนคุยกัน เปิดเสียงวิทยุสื่อสารดังรบกวนผู้อาศัยข้างเคียง 1.2 เสียงรบกวนจากการก่อสร้าง ตึกไม่มีมาตรการควบคุม 2) ฝุ่น - ฝุ่นจากการก่อสร้าง กระทบสุขภาพสร้างความรำคาญ ตึกไม่มีมาตรการควบคุม 3) ควัน - ควันพิษรถ ควันจากการก่อนสร้าง ควันบุหรี่ ตึกไม่มีมาตรการควบคุม 4) ขยะ - ตึกนำขยะมากองไว้หน้าด้านถนน ไม่สร้างห้องหรือสถานที่เก็บขยะอย่างเหมาะสม ส่งกลิ่นรบกวนเมื่อมีฝนตก ทั้งหมด นิติเคยขอร้องหลายครั้งแล้วแต่หลายเดือนแล้วไม่มีการปรับปรุง การก่อสร้างไม่มีระบุระยะเวลาต่อเติมทุบสร้างไปเรื่อยๆ สร้างความเดือดร้อนกับผู้อยู่อาศัยโดยรอบ
Photo:

Date: 2022-06-02

Objective

- Data science is a discipline aimed at data analysis, involving various components, e.g., AI/ML, data preparation, data engineering, data visualization, ML operations, and more.
- Therefore, the objective of this project is to build a practical pipeline and demonstrate diverse, **actionable (end-to-end)** data analysis ideas. Examples include:
  - Example 1: Data Science Pipeline including web scraping → Kafka → workflow control with Airflow → visualization with Power BI
  - Example 2: Big Data Analytics Pipeline including large data ingestion → processing with Spark → visualization through Power BI → storytelling insights

Project criteria

- Each group can have up to 6 members (maximum).
- The project should be a fully functional, end-to-end pipeline that demonstrates practical applications or yields insightful findings. The project must include at least the following 3 components:
  - Component 1: At least one AI/ML component.
  - Component 2: At least one Data Engineering (DE) component.
  - Component 3: At least one visualization (Viz) component that includes either geospatial analysis or graph visualization.
- Including more than three components will enhance the project's depth and interest, with the potential for additional points.

Can we use LLM?

Yes, LLM can be used and considered as AI/ML module.

Since the number of LLM calls can vary, simply calling the API directly is too basic and cannot be counted.

To enhance your work, you should incorporate advanced techniques such as chain-of-thought reasoning, agentic approaches, or loading interesting models (e.g., reasoning models) to make it more compelling.

Web scraping data criteria

- The main data requirement is to analyze the provided dataset from Traffy Fondue. Given that the dataset contains over 700,000 records, you are not required to use all of them but must utilize at least 100,000 records.
- Each group must also use web scraping or an API to add at least 1,000 records from external sources, such as *organization locations, police station locations, PM2.5 status, flooding issues, traffic, etc.*
  - This data collection is a separate task and does not count as part of the DE module.
- Incorporating many data sources will enhance the project's appeal and may earn extra points.

Scoring criteria (10%)

- The scoring criteria will be based on:
- **Completeness:** ensuring all 3 required components + data with web scraping
- **Project interestingness:** evaluated through:
  - Effort (e.g., additional data),
  - Creativity,
  - Execution,
  - Technical quality,
  - and other relevant factors.

Presentation & submission (5%)

- **Presentation Video**
- Upload the video to YouTube and set it to public. The video should be shared in the project channel on Discord by the deadline for other groups to view. The video length is 15 minutes, where the presentation quality will be a key factor in evaluation.
- The video should include:
  - 1. An explanation of the data used, including any additional data sources if applicable.
  - 2. A breakdown of each of the 3 required components, with a diagram illustrating how they interconnect.
  - 3. A demo showcasing interesting results from your data analysis.
- **Other Deliverables**
- Source code, presentation slides (in both PPT and PDF formats).
- Submission should include a link to a Google Drive folder shared on myCourseVille with “viewer” access for anyone.

Important Remarks

- Create group in MCV by Sun 9 Nov 2025
- Submit YouTube and Google Drive link to MyCourseVille by the submission deadline.
- Ensure that your YouTube and Google Drive link are both accessible. If we can’t open it, your work will receive a score of 0.
- Don’t forget to share it in the Discord channel *#project-showroom*, along with a short description of your project.

FAQ