



MONASH
University

L'ORÉAL
MALAYSIA



MUMTEC
MONASH UNIVERSITY MALAYSIA TECH CLUB



**Premier Digital
Tech Institution**



MONASH ANALYTICS CLUB

L'ORÉAL X MONASH DATATHON 2025 HANDBOOK

Table Of Contents

1. EVENT TIMELINE	3
2. PROBLEM STATEMENTS	4
THEME: TrendSpotter	4
THEME: CommentSense	5
3. JUDGING CRITERIA	6
4. PITCH DECK REQUIREMENTS AND GUIDANCE	7
5. SUBMISSION	8
6. DELIVERABLES REQUIREMENTS	9
7. FAQ	10

1. EVENT TIMELINE

Date	Event / Round
August 18, 2025	Official Launch + Start of Registration
August 21, 2025	Kickoff Session (Q&A)
September 5, 2025	End of Registration
September 6, 2025	Submission Portal Opens
September 13, 2025*	Submission Portal Closes
September 24, 2025	Finalist Announcement
October 2, 2025	Finals: Pre-Demo Day at Monash University Malaysia
October 3, 2025	Finals: L'ORÉAL HQ

*(Dates marked with an asterisk are subject to change)**

2. PROBLEM STATEMENTS

THEME: TrendSpotter

Problem Statement
Trends reflect what's currently captivating our community. For L'Oréal, identifying and leveraging them early is crucial for staying relevant, boosting engagement, and driving innovation in beauty.
Deliverables
<ol style="list-style-type: none">1. A functional prototype of an AI-powered solution for trend identification. This prototype should demonstrate the ability to analyze audio snippets, hashtags, and keywords to identify emerging trends.2. Demonstration showcasing the capabilities and benefits of the AI-enhanced trend identification system, potentially including how it identifies key segments participating in a trend (e.g., Gen Z, Millennials) and how it might indicate when a trend starts to decay.3. Recorded presentation
Sub-theme/Challenge
<p>How do we identify a trend?</p> <ul style="list-style-type: none">• Audio snippet used• Hashtags, keywords mentioned• Rate of Change• Multi-platform <p>Which segments are participating?</p> <ul style="list-style-type: none">• Gen Z, Millennials• Sports, Lifestyle, Food, Health, Beauty <p>When is it too late to hop on a trend?</p> <ul style="list-style-type: none">• Optimizing reach and impact• When does trend engagement start to decay?

THEME: CommentSense

Problem Statement
Measuring content effectiveness through Share of Engagement (SoE) metrics like likes, shares, saves, and comments is essential. How do we analyze the quality and relevance of comments, at scale?
Deliverables
<ol style="list-style-type: none">1. A functional prototype of an AI-powered solution for analyzing the quality and relevance of comments at scale. This prototype should demonstrate the ability to measure content effectiveness through Share of Engagement (SoE) metrics like likes, shares, saves, and comments.2. Demonstration showcasing the capabilities and benefits of the AI-enhanced comment analysis system, potentially including a dashboard showing the ratio of 'quality' comments, sentiment breakdowns, categorized comments, and detected spam.3. Recorded presentation
Sub-theme/Challenge
<p>What is our ratio of 'quality' comments?</p> <ul style="list-style-type: none">• Identifying the right KPI to measure efficacy of our messaging <p>What we typically look for:</p> <ul style="list-style-type: none">• Relevance: Does it talk about the product/brand/topic in the post? Is the sentiment good or bad?• Tag/categorise: Can we filter out comments by category? i.e. skincare, fragrance, makeup. <p>What are the common keywords associated with each category?</p> <ul style="list-style-type: none">• Spam Detection

** Upon registration for the L'Oréal x Monash Datathon, students will receive a dataset in the form of a .zip file. This dataset will be general in nature and may be relevant to one or both problem statements. It will be the students' responsibility to determine how best to utilise it in addressing their chosen problem statement.

3. JUDGING CRITERIA

The judges will evaluate teams based on the following criteria.

1. **Innovation & Creativity (30%):**
 - **Problem Approach (15%)** — *How inventive is the team's approach to solving the problem? Does it demonstrate original thinking in addressing data challenges?*
 - **Novelty of Solution (15%)** — *How unique and creative is the proposed AI solution? Does it introduce new approaches or combine existing technologies in innovative ways?*
2. **Technical Execution (20%):**
 - **Design (10%)** — *Quality of solution architecture, data pipeline design, and overall technical approach. Is the solution well-structured and scalable?*
 - **Implementation (10%)** — *The effectiveness of the AI-driven prototype in addressing the problem, including code quality, performance, and scalability.*
3. **Functionality (30%):**
 - **Solving Problem Statement (15%)** — *How well does the prototype solve the problem statement's requirements?*
 - **Usability & Insightfulness (15%)** — *Can stakeholders (e.g., marketing, product) take action based on the result?*
4. **Presentation & Documentation (20%):**
 - **Quality & Completeness of Presentation (10%)** — *Clarity, organization, and effectiveness of the team's presentation. Does it clearly explain the problem, approach, and results?*
 - **Quality & Completeness of Demo (10%)** — *Clarity and thoroughness of the recorded presentation and documentation. It is recommended to include an overview of the prototype in the presentation*

Total Points: 100%

- Please bear in mind the relevance of the project regarding the division and market needs and the general appearance of the presentation.
- Where two (2) or more Teams scored the same number of points, the Organisers will have the full discretion to decide on which Team is to be one of the Winners and the Prize that is to be allocated accordingly. In the event that there are several submissions with the same score, the jury will be free to decide which project they want.
- Please ensure your submission complies with the official competition rules. These include guidelines on originality, integrity, and prohibited actions. The full rules are provided in the [Competition Rules & Regulations](#) document.

4. PITCH DECK REQUIREMENTS AND GUIDANCE

1. Project Overview

- **Problem/Opportunity Statement:** Clearly state the problem you are solving or the opportunity you are addressing.
- **Existing Solutions:** Discuss existing solutions, deployments, or research related to your project and highlight their issues or limitations.

2. Project Benefits

- **Target Audience:** Define your project's target audience.
- **Advantages:** Explain the advantages and unique value proposition of your project compared to existing methods (as discussed in your overview).

3. Project Deliverables

- **Final Product:** Describe the final deliverable (e.g., an app, portal, script, API).
- **Deployment:** Outline the requirements for deploying your project (e.g., "Deployment requires X, Y, and Z").

4. Development Methodology

- **Tools and Technologies:**
 - List the software and hardware tools used in the development process.
 - **Do NOT** reference project management tools like Trello, Kanban boards, WhatsApp, or GitHub.
- **Licensing and Costs:**
 - If applicable, detail any licensing or associated costs.
 - Provide examples like hosting on AWS or a subscription to a specific API.

5. Future Plans and Possibilities

- **Expansion Possibilities:**
 - Discuss potential expansion possibilities or future features for the software/project if given more time.
 - Highlight any limitations or scopes that were set for the current project.
- **Overcoming Limitations:**
 - Explain what limitations could be overcome with additional budget or time.

5. SUBMISSION

Submission will open on **6th September 2025** and must be completed by **13th September 2025, 11:59 PM MYT**.

The submission link will be sent to the registered team's email address.

Each team member will receive the submission link, but please coordinate internally to ensure that **only one designated member** submits on behalf of the team.

If multiple submissions are received and the team does not clearly indicate which one should be judged, **the latest submission will be taken as the final version for evaluation**.

Your submission must include:

- Link to the Pitch Deck (PDF or PPT)
- Link to your presentation video (maximum 5 minutes):
 - Explaining the problem addressed, the AI solution provided, and its potential impact on L'Oréal's innovation in beauty.
 - Demonstration showcasing the capabilities and benefits of the AI-enhanced prototype.
 - Data Pipeline & Code Coverage (where applicable) should show how raw data is processed and integrated into the AI solution with relevant code coverage to demonstrate completeness.
- Link to the GitHub/Prototype link as a Zip file

6. DELIVERABLES REQUIREMENTS

- During the Competition, Participants may upload any documents or other material (such as texts, presentations, videos, etc.) relating to their project (the “Deliverable”). All the Deliverables must be done in **English**.
- All Entries and Deliverables submitted to the Organisers at any stage of the Competition must strictly comply with the guidelines shared throughout the Competition Period and adhere to the intellectual property provisions as contained in this Official Rules.
- Upon submission of any Deliverables at any stage of the Competition, each Participant guarantees to the Organisers that they are the legal owner and/or legitimate holder of the intellectual property rights pertaining to all or any part of the creations contained within the Deliverables. Furthermore, each Participant affirms that they have secured all necessary rights and permissions for any pre-existing creations and any elements of the Deliverables for which they do not hold the relevant rights.
- Should any Entry of the Deliverables incorporate any intellectual property right belonging to a party who is not a Member of the Team, whether registered or not (e.g., patents, trade secrets, know-how), Participants must explicitly notify the Organisers in writing before using or incorporating such intellectual property right in the Deliverables and/or Entry.
- The Organisers reserve the right to disqualify an Entry that contains any intellectual property rights belonging to a party who is not a Member of the Team.
- Each Participant guarantees (i) that their contribution is original and unprecedented, and (ii) that it does not infringe upon the intellectual property rights of any third party.
- L'Oréal will not be held liable for any infringement of the above provisions by the Participants. Each Participant shall indemnify The Organisers against any disturbance, action, claim, opposition, demand, or eviction attempt from any third party in connection with any creation within the Deliverables (including Intellectual and Industrial Property Rights).

7. FAQ

Case Study & Technical Questions

Q1: Are there any preferred tools, libraries, or frameworks we should use?

You are free to use any tools, frameworks, or programming languages you are comfortable with. Python is generally preferred, and libraries like scikit-learn, pandas, TensorFlow, PyTorch, and Hugging Face are all allowed.

Q2: Can we use pre-trained models like BERT or GPT from Hugging Face?

Yes, you may use pre-trained models if they enhance your solution. Be sure to clearly explain their use in your final submission.

Q3: Is the CSV provided just a sample? Can we expect more data, or should we augment with external datasets?

The provided dataset is the primary source for your analysis. However, if you believe external datasets will improve your solution (e.g., trend detection), you are welcome to use them. Please cite your sources and explain how they contribute.

Q4: Can ensemble models be used?

Yes, ensemble methods are allowed. You're encouraged to use techniques that improve your model's robustness, as long as the approach is well-justified and reproducible.

Q5: How should we define “emerging” vs. “established” trends quantitatively?

We leave the definition open to interpretation to encourage creativity. However, we expect your submission to include a clear rationale and quantitative criteria for how you distinguish between these two trend categories.

Q6: Is there a specific output format we should follow?

You will be provided with a submission guideline. In general, we expect a presentation deck, and optionally, a notebook or script in your Zip file that explains your methodology. A dashboard or lightweight prototype is welcome but not mandatory unless stated otherwise.

Q7: Are there specific stakeholders we should tailor our solution towards? (e.g., marketing, product innovation)

Yes, your recommendations should ideally align with business functions such as marketing, product development, or trend forecasting. You are encouraged to state clearly who your intended stakeholder is.

Q8: How should we handle multilingual comments? Should we translate them or use multilingual models?

Either approach is valid. You can translate the comments first or use multilingual pre-trained models like XLM-R. Do mention any trade-offs (e.g., speed vs. accuracy) in your approach.

Q9: Some comments contain odd characters like “öÿ”. Should we discard or fix them?

These are likely encoding issues from emojis or special characters. We recommend attempting to clean or decode them where possible. If unresolvable, document your decision to discard and justify it.

Q10: Are the comments directly tied to videos via a videoID?

Yes, each comment includes a videoID, which you can use to group or analyze comments by video context if it supports your insights.

Q11: Should sentiment be scored on a scale (1-5) or is a basic positive/neutral/negative classification enough?

The baseline expectation is to classify sentiments into positive, neutral, or negative categories. However, if you believe a more granular scale adds value to your analysis, feel free to implement it and justify the approach.

Q12: Can we assume that higher like counts indicate better comment quality? Should we consider other factors?

While like count is a useful proxy for engagement, it's not always sufficient. You may want to factor in comment length, topic relevance, or sentiment strength. We encourage critical evaluation of what defines "quality" for your use case.

Q13: How will our work be evaluated—based on accuracy, creativity, or business relevance?

Your solution will be judged based on a mix of:

- Technical accuracy & robustness
- Innovation & creativity
- Business relevance & stakeholder alignment

Please refer to the judging criteria for more information.

Q14: Is a dashboard or app prototype required, or is backend analysis enough?

A working prototype (e.g., dashboard or app) is optional but can enhance your submission. The core focus is on actionable insights and strong analytical thinking. Prioritize quality of analysis, then complement with a visual/demo element if time permits.

Event Logistics

Q15: Will mentors be provided during the datathon?

For preliminary rounds, there will be **no mentors available for guidance**. However, **mentors** will be available for finalists only during the Pre-Demo Day to guide them in both technical aspects and business understanding.

Q16: Will food be provided?

Absolutely. Meals and refreshments will be provided during the Pre-Demo day and Final Event days to keep everyone energized.

Q17: Will accommodation be provided for students traveling from other states?

No, accommodation will not be provided; however, we will share a list of recommended places for students from other states.