Structured Approach for Milestone 1: Problem Identification

This document outlines a structured approach for our team to successfully complete Milestone 1: Problem Identification, drawing upon the key concepts and frameworks discussed in the syllabus and recent workshops. The core objective of this milestone is to transition from a broad understanding of a problem to a clearly defined, actionable research question that can be addressed through data science.

Phase 1: Empathize and Diverge (Understanding the Problem Space)

This initial phase is about broad exploration and deep understanding of the problem. It aligns with the "Empathize" stage of Design Thinking and the concept of "Divergent Thinking" discussed in your workshops. The goal is to cast a wide net, gather as much information as possible, and truly understand the problem from various perspectives, especially those of the affected individuals or stakeholders.

Step 1.1: Personal Problem Identification and Brainstorming

Begin by identifying problems based on your personal experiences. As mentioned in the syllabus, Milestone 1 requires "A problem-statement based on your personal experiences." This is your starting point for divergent thinking. Each team member should individually brainstorm problems they have encountered or observed in their daily lives that they believe could potentially be addressed with data science. Think broadly and do not self-censor at this stage. Consider problems within domains of interest such as Economic Inclusion, Education, or Healthcare, as suggested in the syllabus.

Step 1.2: Initial Domain Study and Background Review

Once individual problems are identified, the team should collectively select a few promising areas for an initial domain study. This aligns with the deliverable of "A thorough background review of your research domain in the `0_domain_research` folder of your repository." The purpose of this step is to gain a foundational understanding of the chosen problem areas. This involves:

- **Literature Review:** Search for existing research, reports, articles, and data related to the identified problems. Understand what has already been done, what solutions have been attempted, and what gaps still exist.

- **Stakeholder Identification:** Identify potential stakeholders who are affected by or involved with the problem. These could be individuals, communities, organizations, or even systems. Understanding their perspectives is crucial for empathizing with the problem.

- **Domain Expertise Exploration:** Begin to appreciate the importance of domain expertise. While you don't need to be an expert, understanding the nuances of the domain will help in framing a relevant and impactful research question. Consider if there are experts you could consult (though this might be more relevant in later stages, initial identification is useful).

Step 1.3: Systems Thinking Application (Understanding the "Iceberg")

As emphasized in the June 4th meeting, apply systems thinking to look beyond the surface-level symptoms of the problem. This is about understanding the underlying structures and patterns that contribute to the problem, rather than just the visible "tip of the iceberg." This step directly supports the deliverable of "A summary of our group's understanding of the problem domain (Bonus points for using systems thinking!)." Ask yourselves:

- What are the observable events or symptoms of the problem?

- What patterns or trends can be identified over time related to these events?

- What underlying structures (policies, power dynamics, cultural norms, infrastructure) contribute to these patterns?

- What mental models or assumptions are held by various stakeholders that perpetuate the problem?

By delving into these deeper layers, our team can gain a more holistic understanding of the problem, which is essential for defining a truly actionable research question.

Phase 2: Define and Converge (Framing the Research Question)

This phase is about narrowing down your broad understanding into a specific, actionable problem statement and then formulating a precise research question. This aligns with the "Define" stage of Design Thinking and the concept of "Convergent Thinking."

Step 2.1: Problem Statement Formulation

Based on your empathic understanding and domain study, formulate a concise problem statement. This should clearly articulate the problem our team intends to address. Remember the syllabus requirement for "A problem-statement based on your personal experiences" to be written in the `README.md`. Your problem statement should be:

- **User-Centered:** Clearly identify who is experiencing the problem.

- **Problem-Focused:** Describe the core issue or challenge.

- **Contextualized:** Provide enough background to understand the problem's setting.

Step 2.2: Identifying Constraints and Scope

Before framing the research question, explicitly identify our group's constraints. Constraints can be voluntary (chosen by the group to ensure project completion) or involuntary (external limitations). These constraints will help us scope our project realistically. Consider:

- **Time:** The project timeline (Milestone 1 due June 16, overall project end August 25).

- **Available Data:** What data is realistically accessible? (This will be a major factor, as highlighted in the June 4th meeting regarding data costs).

- **Domain Expertise:** Our team's collective knowledge and access to external experts.

- **Technical Capabilities:** The skills and tools our team possesses.

- **Resources:** Any other limitations on budget, software, or hardware.

These constraints will help you narrow down the problem space to something manageable and achievable within the project's scope. This directly relates to the syllabus point of identifying problems "accessible within our constraints."

Step 2.3: Framing an Actionable Research Question

This is the most critical deliverable for Milestone 1: "An actionable research question informed by the realities of your problem domain and your constraints." A good research question, as emphasized in the June 5th meeting, is the "bullseye" of our project. It should be:

- **Specific:** Clearly defined, avoiding ambiguity. (Recall the "vase vs. experiencing flowers" exercise).

- **Measurable:** Can be answered with data.

- **Achievable:** Feasible given our constraints and resources.

- **Relevant:** Addresses a significant aspect of the problem and has potential impact.

- **Time-bound:** (Implicitly, within the project timeline).

Our research question should connect:

- **What you are measuring:** The variables or indicators.

- **The problem indicators are related to:** The core issue.

- **Who it affects:** The target population or stakeholders.

Example of refining a research question (from June 3rd meeting):

- **Too Broad:** "How does social media impact mental health in teenagers?"

- **More Specific/Actionable:** "How does social media use over 3 hours a day impact mental health in teenagers at this particular age?"

Work collaboratively to draft and refine our research question. It's an iterative process, so expect to go through several versions. Ensure it can be answered using data science methodologies.

Phase 3: Documentation and Collaboration

Throughout Milestone 1, consistent documentation and effective collaboration are paramount. These are not separate steps but ongoing activities that support all phases.

Step 3.1: Maintain Planning Documents

As per the syllabus, continue to "Maintain our planning documents (group norms, learning goals, constraints, a communication plan, …)." These documents are living resources that should be updated as our understanding of the project evolves.

Step 3.2: GitHub Repository Management

Ensure our GitHub repository is well-maintained and well-documented. The syllabus explicitly states that deliverables will be checked by directly reviewing our group's repository. Specifically, for Milestone 1:

- Create the `0_domain_research` folder and place your thorough background review there.

- Write your problem statement, summary of problem domain, and actionable research question in the `README.md` file.

Step 3.3: Retrospective and Survey

Complete "A retrospective for this milestone" and "The milestone survey." These are crucial for reflecting on our team's process, identifying areas for improvement, and demonstrating our learning journey.

Step 3.4: Git Tagging

Create "A labeled Git tag for this milestone created before the deadline." This ensures that our work is clearly marked for assessment.

By following this structured approach, our team can effectively navigate Milestone 1, identify a meaningful problem, and frame a clear, actionable research question, setting a strong foundation for the subsequent phases of our data science project.