

Problem Definition & Topic Justification

Checkpoint 2

This checkpoint builds directly on your Proposal Memo. You'll refine your project into a clear, measurable problem statement and demonstrate feasibility through defined KPIs, data planning, and early analytical steps. This memo ensures your project remains focused, data-driven, and ready for execution

Overview.....	1
Group Work Option.....	1
Assignment Objectives.....	1
What to Include in Your Memo.....	2
Submission Details.....	2
Grading & Rubric (10 points total).....	3
Additional Notes.....	3

Overview

Submit a professional memo (~1-2 pages) that refines the problem statement, identifies clear KPIs, confirms data feasibility, and outlines early analytical steps (EDA, preprocessing, preliminary modeling). Write as if communicating to a sports stakeholder.

Group Work Option

You may work in groups of up to 3 students, but each student must maintain their own GitHub repository. Collaboration on shared data and topic is allowed, but your individual memo and repository submission are required.

Assignment Objectives

By completing this checkpoint, you will:

1. Refine your topic into a well-defined, measurable problem statement.
2. Identify and justify KPIs that align with your stakeholder or decision context.
3. Demonstrate the feasibility of your dataset(s) and defined scope.
4. Outline EDA, preprocessing, and modeling plans tied to your KPIs.
5. Practice writing a concise, professional stakeholder memo.

What to Include in Your Checkpoint

Section	Description	Guiding Questions
Project Title & Updated Topic Summary	Provide a working title and concise 1-2 sentence overview of your refined topic.	How has your topic evolved or narrowed since Checkpoint 1?
Problem Definition & Justification	Define your research question and explain why it matters to a real stakeholder decision.	Why does this question matter? Who will use the insights?
Key Metrics (KPIs)	List 1-2 measurable variables that align with success or performance goals.	Which metrics best represent your objective?
Unit of Analysis & Scope	Define your observation level (play, game, player-season, etc.) and boundaries (season, league, filters).	What counts as one observation, and what will you include/exclude?
Data Sources & Access Plan	Identify dataset(s) and how you'll obtain or prepare them.	Is data publicly accessible and sufficient for analysis?
Mini Literature Scan	Summarize findings from ≥ 3 relevant sources supporting your problem design.	What similar research exists? How will yours add value?
Exploratory & Preprocessing Plan	Outline your early analysis approach and cleaning plan.	How will you inspect and prepare your data (EDA checks, missingness, encoding, scaling)?
Modeling Direction (Preliminary)	Name 1-2 analytical or modeling paths you might pursue.	What models best fit your question and KPIs? How will you evaluate them?
Risks & Ethical Considerations	Identify potential biases, data access issues, or privacy concerns and mitigations.	What risks exist and how will you manage them?
Next Steps	Summarize immediate tasks for the next week.	What actions will you take before Checkpoint 3?

Submission Details

Requirement	Details
Document Type	Google Doc (Professional Memo Format)
Submission	Share your Google Doc link on Classroom and upload a PDF copy to /reports in your GitHub repo.
Format Tips	Include your name(s), project title, and checkpoint heading at the top of your memo.

Grading & Rubric (10 points total)

Your Checkpoint 2 will be evaluated on the following criteria:

Criteria	Description	Points
Clarity of Problem Definition	Question and KPIs are precise, measurable, and relevant.	2
Data Feasibility & Scope	Data and scope are clear, obtainable, and realistic.	2
Relevance & Justification	Stakeholder context and significance clearly defined.	
Literature & Method Alignment	Literature scan informs methods and adds credibility.	2
Analytical Planning	EDA, preprocessing, and model ideas are logical and well-linked to objectives.	2
Professional Structure & Writing	Memo is polished, organized, concise, and professional.	2

Additional Notes

- *Keep it professional and readable, aim for clarity over length.*
- *Write as if addressing a coach, manager, or executive in a sports context.*
- *Focus on an actionable problem that data can realistically address.*
- *State assumptions clearly and note any data limitations.*
- *Maintain consistency with your GitHub project organization.*

Value on the Court: Assessing NBA Player Efficiency across Salary Tiers Overview The following project examines how NBA player efficiency differs across salary tiers in search of which of the salary groups provides the most on-court value per dollar spent. This analysis has been narrowed down from a general performance-based one to a more focused one, comparing efficiency metrics across well-defined salary bands.

Do NBA players in the lower tiers of salaries bring similar or greater efficiency than their higher salaried counterparts?

Justification: This is important for team executives, salary cap analysts, and player development staff who need to optimize roster construction within financial constraints. The intuition gained can assist in contract negotiations, trade decisions, and player development priorities.

KPI

PER: Player Efficiency Rating sums up a player's total impact per minute through a composite metric.

Win Shares per 48 Minutes (WS/48): A measure of a player's contribution to team wins, adjusted for playing time.

Unit of Analysis: Individual player-season

Scope:

Seasons: 2018–2023

League: NBA only

Filters: minimum 500 minutes played per season to ensure meaningful sample size

Mini Literature Scene

Berri & Schmidt 2006 - Found that salaries are generally poorly correlated with actual performance, suggesting inefficiencies in player valuation.

Goldsberry 2019 - Highlighted the growing importance of advanced metrics like WS/48 and PER in assessing player impact beyond traditional box score statistics.

Sportrac Analytics Blog 2022: Salary-to-efficiency ratios highlighted the trends in overpaid and underpaid players.

EDA Tasks:

Check for missing salary or efficiency data

Visualize distributions of PER and WS/48 across salary tiers

Preprocessing:

Encode categorical variables (e.g., position, team)

Even out salaries to account for inflation or salary cap changes.

Scale efficiency metrics for regression modeling

Models

Linear Regression: To understand the correlation between salary and efficiency metrics.

Clustering (e.g., K-Means): Identify natural clusters of players given salary and performance.

Bias Risk: Salaries might reflect the market dynamics, not just performance. Examples include veteran premiums and rookie scale contracts.

One-Way ANOVA: Control for age, experience, and position

Privacy: All data is public; no personal or sensitive information is used.