

Course Project Rubric Handout

Spring 2026

Instructor: Dr. Sarangan (Ravi) Ravichandran, Ph.D.

This rubric describes how the course project will be evaluated across each milestone. Total = 100 points.

Project Structure and Due Date Conventions

Project structure (two major milestones):

- Midterm milestone (20%): EDA + problem framing + code demo and reasoning (Project Midterm Presentation).
- Final milestone (30%): modeling + evaluation + interpretation + limitations (Final Demo & Discussion).
- Milestone due dates in the project handout are labeled by class meeting (e.g., "Class 3"). These correspond directly to the week numbers in the 16-week course schedule (e.g., Class 3 = Week 3).
- The course spans 16 calendar weeks with 15 class meetings due to a holiday/no-class week.
- All deadlines labeled "Class X" refer to the Xth class meeting.

Grading Summary

Milestone	Due	Points
1) Self-Introduction & Learning Goals	Class Meeting 1	5
2) Project Proposal	Class Meeting 3	10
3) Progress Check-In #1 (Preliminary EDA Notebook)	Class Meeting 7	10
4) Midterm Presentation (EDA-focused)	Class Meeting 8	20
5) Progress Check-In #2	Class Meeting 13	10
6) Final Demo & Discussion (Modeling-focused)	2 days before the last class; check BB for details	30
7) GitHub Repository Quality	2 days before the last class; check BB for details	15

Performance Levels

Exemplary (E): Meets all requirements and demonstrates strong clarity, correctness, and insight.

Proficient (P): Meets most requirements; minor gaps that do not substantially affect correctness.

Developing (D): Partially meets requirements; multiple gaps or unclear reasoning/communication.

Needs Work (N): Missing major components, incorrect, not reproducible, or not aligned to the prompt.

1) Self-Introduction & Learning Goals (5 pts)

Criteria	Pts	Exemplary (E)	Proficient (P)	Developing (D)	Needs Work (N)
Completeness (background, interests, solo/team preference)	2	All requested items included and clearly stated.	One minor item unclear or slightly underdeveloped.	Multiple items vague or only partially addressed.	Missing key required items.
Learning goals and skills to gain/improve	2	Specific goals and concrete skills to build; clear alignment to course.	Goals present but somewhat general; partial alignment.	Goals are generic or weakly connected to course.	Goals/skills not provided or not meaningful.
Writing quality and length (150-200 words)	1	Within range; clear, professional, and well organized.	Minor length or clarity issues; still readable.	Hard to follow, repetitive, or noticeably off-length.	Very unclear, not within range, or not submitted.

2) Project Proposal (10 pts)

Criteria	Pts	Exemplary (E)	Proficient (P)	Developing (D)	Needs Work (N)
Dataset selection (name, source/URL, rationale)	3	Dataset is appropriate; source/URL provided; strong rationale.	Appropriate dataset; source/URL provided; rationale adequate.	Dataset or rationale weak/unclear; source may be incomplete.	Dataset/source missing or not appropriate.
Problem statement / research question	3	Clear, focused, and testable question with expected outcome.	Mostly clear and testable; minor scope issues.	Too broad or vague; hard to evaluate.	Missing or not testable.
Planned methods (full-semester plan)	3	Describes an end-to-end approach across the semester (EDA → modeling → evaluation → interpretation), with a clear plan and justification.	End-to-end plan is present and mostly appropriate; limited justification or missing one component.	Plan is partially complete, mismatched, or unclear (e.g., missing evaluation or interpretation).	Methods are missing or not aligned to the project/course.

Feasibility and scope	1	Well scoped for time and tools; risks acknowledged.	Likely feasible; minor scope concerns.	Over/under-scoped; feasibility questionable.	Not feasible within course constraints.
-----------------------	---	---	--	--	---

3) Progress Check-In #1 (Preliminary EDA Notebook) (10 pts)

Criteria	Pts	Exemplary (E)	Proficient (P)	Developing (D)	Needs Work (N)
Reproducible setup (data loaded; notebook runs)	2	Notebook runs end-to-end without errors; environment clearly described.	Minor run issues; easy to fix; mostly reproducible.	Multiple run issues; reproducibility unclear.	Not runnable or missing notebook.
Initial cleaning / preprocessing	3	Correct cleaning steps with brief justification.	Cleaning mostly correct; minor gaps.	Minimal or unclear cleaning; errors possible.	No meaningful cleaning or incorrect approach.
EDA visualizations (2-3) with labels	3	Plots are correct, well labeled, and informative.	Plots mostly correct; minor labeling/interpretation gaps.	Plots weak, confusing, or limited insight.	Missing or incorrect plots.
Summary statistic(s) + short progress note	2	Correct stats plus clear progress update and next steps.	Stats provided; progress note somewhat clear.	Stats or note incomplete/vague.	Missing stats and/or progress note.

Data cleaning means preparing raw data so it is accurate, consistent, and ready for analysis. Raw datasets often contain missing values, formatting issues, or inconsistencies that must be addressed before analysis or model building.

Data cleaning can include:

- Organizing the dataset into a clear structure (sometimes called *tidy data*)
- Handling missing values (for example, removing or filling NA values)
- Fixing errors or inconsistent entries
- Correcting data types (such as converting text to numbers or dates)
- Removing duplicate or irrelevant records

For this course, when you perform **initial cleaning/preprocessing**, I am looking for:

- ✓ Reasonable steps that prepare the dataset for analysis
- ✓ A short explanation of what you cleaned and why

There is usually no single “correct” way to clean data, but your approach should be logical and clearly explained.

4) Midterm Presentation (EDA-focused) (20 pts)

Criteria	Pts	Exemplary (E)	Proficient (P)	Developing (D)	Needs Work (N)
Structure, clarity, and timing (5-7 minutes)	4	Well organized narrative; within time; slides/demo support message.	Mostly organized; slight timing or clarity issues.	Disorganized or rushed; time management issues.	Hard to follow or significantly off time.
Dataset overview and context	4	Explains dataset, key variables, and context clearly.	Overview mostly clear; minor gaps in context/variables.	Thin context or unclear variables.	Missing/unclear overview.
EDA quality (≥ 3 plots + summary)	6	Correct plots with strong labeling and meaningful summary insights.	Plots mostly correct; limited insights or minor issues.	Plots or summary weak; limited interpretation.	Missing/incorrect EDA outputs.
Interpretation and early insights ('so what')	4	Clear takeaways tied to the question; thoughtful hypotheses/next steps.	Some takeaways; connection to question mostly clear.	Mostly descriptive; limited insight.	No meaningful interpretation.
Reproducibility and next steps	2	Demo runs; concrete next steps and plan.	Mostly reproducible; next steps stated.	Reproducibility unclear; next steps vague.	Not reproducible; no plan.

5) Progress Check-In #2 (10 pts)

Criteria	Pts	Exemplary (E)	Proficient (P)	Developing (D)	Needs Work (N)
Progress since midterm	4	Clear and substantial progress with new results.	Moderate progress with some new results.	Limited progress; small changes.	Minimal/no progress.
Added methods beyond basic EDA	3	Appropriate additional technique(s)	Technique applied with minor issues or	Attempted technique but	No additional methods.

		applied correctly and explained.	limited explanation.	weak/incorrect or unclear.	
Notebook quality (readable, runnable)	2	Clean, organized, reproducible notebook.	Minor organization/run issues.	Multiple clarity or run issues.	Not runnable or missing.
Written update (1-2 paragraphs)	1	Clear status, challenges, and plan.	Mostly clear but brief.	Vague or incomplete.	Missing.

6) Final Demo & Discussion (Modeling-focused) (30 pts)

Criteria	Pts	Exemplary (E)	Proficient (P)	Developing (D)	Needs Work (N)
Modeling approach and justification	8	Appropriate model(s); clear rationale; aligns with question and data.	Model(s) mostly appropriate; rationale adequate.	Model choice questionable or poorly justified.	Missing or incorrect modeling approach.
Validation and evaluation	8	Correct validation (train/test or CV); appropriate metrics; interpreted correctly.	Mostly correct validation/metrics; minor interpretation gaps.	Partial/unclear validation; metrics weak or misused.	No meaningful validation or incorrect evaluation.
Results interpretation + limitations + next steps	6	Thoughtful interpretation, limitations, and concrete future work.	Interpretation and limitations present; somewhat general.	Superficial discussion; limited limitations.	Missing discussion/limitations.
Presentation quality (visuals, narrative)	5	Professional delivery; clear visuals; strong story.	Mostly clear; minor visual/story issues.	Uneven clarity; cluttered or confusing visuals.	Hard to follow.
Live repo walkthrough and Q&A readiness	3	Clear walkthrough of repo/notebook; answers questions well.	Walkthrough present; some Q&A gaps.	Minimal walkthrough; struggles with questions.	No walkthrough; cannot explain work.

7) GitHub Repository Quality (15 pts)

Criteria	Pts	Exemplary (E)	Proficient (P)	Developing (D)	Needs Work (N)
Repo structure matches required layout	4	Follows required structure (notebooks/, data/, results/, src/, README.md).	Mostly follows structure; minor deviations.	Some key folders missing or confusing structure.	Structure missing/incorrect.
README completeness (required items)	5	Complete and easy to follow; includes goal, data source, methods, how to run, results summary, and references.	Mostly complete; minor omissions.	Several omissions; unclear instructions.	Missing or very weak README.
Reproducibility (runs without modification)	4	Notebook/scripts run without changes in Colab or Jupyter; dependencies stated.	Minor fixes needed; mostly reproducible.	Multiple fixes needed; dependencies unclear.	Not reproducible.
Code/notebook hygiene	2	Readable, commented, consistent naming; minimal clutter.	Some clarity issues.	Messy/unstructured; hard to read.	Very poor quality.