Report Writing & Documentation

Lara

- Led the overall structure of the final report
- Wrote sections:
    - Introduction
    - Dataset Description
    - Methodology overview
    - EDA insights
    - Discussion + Implications

Vivek

- Authored core technical sections of the report
- Wrote sections:
    - Preprocessing Pipeline
    - Feature Engineering
    - Model Architectures (Regression + Classification)
    - Results interpretation (regression + classification tables)
- Assisted in preparing model evaluation summaries and Appendix A tables.
- Integrated references, formatting, polishing, and narrative cohesion.

Swathi

- Contributed to the "Limitations," "Future Work," and parts of "Conclusion"
- Helped review and edit the EDA and Feature Importance sections
- Ensured alignment between report and slide visuals/models.

Sarah

- Contributed to Dataset Limitations, Methodology refinement, and Discussion wording
- Assisted in proofreading the entire report for clarity and consistency
- Ensured terminology alignment with class expectations and DS3000 rubric.

Slide Deck Creation

Swathi

- Created slides for:
    - Problem Statement
    - Dataset Overview
    - Preprocessing Pipeline
    - EDA (visuals, summary text)
- Ensured that visuals matched the report's figures.

Sarah

- Designed slide layout, colour scheme, and formatting
- Created slides for:
    - Model Overview (Regression + Classification)
    - Results slides (metrics, graphs, confusion matrix, ROC curve)

    - Feature importance + limitations + takeaways
- Refined transitions and added speaking notes.

Lara

- Inserted narrative framing, intro/outro, and takeaway messaging
- Ensured the demo + Q&A section matched group talking points.

Vivek

- Verified numerical values in results slides
- Added model diagrams and technical notes to ensure accuracy.

Coding & Model Development

Shared Responsibilities

- Collaboratively implemented the full machine learning workflow:
    - Data loading, cleaning, and preprocessing
    - One-hot encoding + normalization pipeline
    - Train/validation/test splitting
    - Regression models (Linear, Ridge, Lasso, Random Forest)
    - Classification models (Logistic Regression, Random Forest, KNN, Decision Tree)
    - Feature importance extraction
    - Plotting EDA visuals and confusion matrix / ROC curve

- Each member contributed to debugging and testing portions of the Jupyter Notebook.
- Lara: data cleaning, EDA visualization scripts
- Vivek: model training pipelines + evaluation helper functions
- Swathi: classification model tuning + plotting confusion/ROC
- Sarah: preprocessing checks, categorical encoding, feature engineering tests

All four reviewed and validated each other's code before final submission.

Presentation Delivery

Evenly Shared Contributions

- Each member prepared and delivered their portion of the verbal presentation
- Everyone practiced timing, flow, and Q&A readiness
- Speaking roles were organized by section:
    - Lara: Motivation + Problem + Dataset
    - Swathi: Methodology + Models + EDA
    - Vivek: Results + Preprocessing + Insights
    - Sarah: Feature Importance + Limitations + Takeaways