Project 4

Start Assignment

- Due Mar 4 by 12:59am
- Points 100
- Submitting a text entry box or a website url

Requirements

Data Model Implementation (25 points)

- A Python script initializes, trains, and evaluates a model (10 points)

- The data is cleaned, normalized, and standardized prior to modeling (5 points)

- The model utilizes data retrieved from SQL or Spark (5 points)

- The model demonstrates meaningful predictive power at least 75% classification accuracy or 0.80 R-squared. (5 points)

Data Model Optimization (25 points)

- The model optimization and evaluation process showing iterative changes made to the model and the resulting changes in model performance is documented in either a CSV/Excel table or in the Python script itself (15 points)

- Overall model performance is printed or displayed at the end of the script (10 points)

GitHub Documentation (25 points)

- GitHub repository is free of unnecessary files and folders and has an appropriate .gitignore in use (10 points)

- The README is customized as a polished presentation of the content of the project (15 points)

Group Presentation (25 points)

- All group members speak during the presentation. (5 points)

- Content, transitions, and conclusions flow smoothly within any time restrictions. (5 points)

- The content is relevant to the project. (10 points)

- The presentation maintains audience interest. (5 points)

This project will be evaluated against the requirements and assigned a grade according to the following table:

Grade	Points
A (+/-)	90+
B (+/-)	80−89
C (+/-)	70−79
D (+/-)	60−69
F (+/-)	< 60

Submission

To submit your project, click Submit, and then provide the URL of your GitHub repository for grading.