

Rubric: Can You Predict Which Restaurants Will Survive?

Purpose: Use real-world Yelp data to explore whether online reviews impact restaurant survival in Philadelphia. Apply natural language processing and decision tree modeling to answer this question.

Specification Category	Specification Details
Clean & Prepare Data	<ul style="list-style-type: none"><input type="checkbox"/> Filter to Philadelphia restaurants only.<input type="checkbox"/> Merge review and business datasets.<input type="checkbox"/> Aggregate sentiment scores per restaurant.
Perform Sentiment Analysis	<ul style="list-style-type: none"><input type="checkbox"/> Use VADER to compute compound sentiment scores for each review.<input type="checkbox"/> Average sentiment per restaurant.
Modeling	<ul style="list-style-type: none"><input type="checkbox"/> Build a decision tree to predict whether a restaurant is open (1) or closed (0).<input type="checkbox"/> Split the data into training and testing sets.<input type="checkbox"/> Evaluate model performance (ROC AUC, accuracy, precision, recall).
Interpret Results	<ul style="list-style-type: none"><input type="checkbox"/> What features are most important?<input type="checkbox"/> How well does the model perform?

	<input type="checkbox"/> Do online reviews actually help predict restaurant success?
Deliverable	<input type="checkbox"/> Jupyter Notebook or R Script with all code and explanations. <input type="checkbox"/> Short report summarizing your findings, insights, and suggestions.

Evaluation: The criteria below is required to either receive a Complete or Incomplete for this project:

Specification Category	Complete	Incomplete
Clean & Prepare Data	Data is well-cleaned, clearly filtered, and prepped	Major gaps or uncleaned data
Perform Sentiment Analysis	VADER applied correctly, logic is well-explained	Incorrect or missing analysis
Modeling	Model built, evaluated thoroughly, with solid justification	Incorrect or absent model
Interpret Results	Insightful analysis, connects results to real-world impact	Lacks interpretation or incorrect conclusions
Deliverable	Notebook/report is polished, readable, and well-documented	Poor structure, unclear or hard to follow