Case Study

NYC Restaurant Inspection Risk Classification

Dataset: NYC Open Data – Restaurant Inspections (2019 sample)

Download Link: Click here

Tasks:

- Load & Filter: Select columns, drop entries without risk_category.
- Text & Categorical Prep: Clean violation_code & action, encode borough & month.
- Modeling: Predict risk_category (Low/Moderate/High) with multiclass classifier; handle imbalance.
- Evaluation & Visualization: Report accuracy, macro-F1, confusion matrix; map high-risk by borough.
- Takeaways: Identify violation codes predicting high risk.

Bonus (Generative AI):

Prompt GPT to draft health department alerts for rising high-risk areas.

Deliverables:

- Jupyter Notebook with code and analysis.
- 5–7-slide Deck summarizing approach and results.
- README with setup & execution instructions.