Evaluation of

Author Name

Table of contents

Preface

You are a data scientist for a mid-sized business, in a small group of 3-4 data scientists. You've been tasked with creating a report evaluating a scenario for your business. Your colleagues will also be evaluating the same scenario, and your reports will be used in aggregate to determine a consensus (or lack thereof) on the company's action. The reports will also be used to inform downsizing that is rumored to be coming - you want to ensure your report is better than your peers so that you aren't as easy to cut.

You may talk to your peers who are assigned the same scenario, but you do not want to collaborate too closely, lest you both become targets of the rumored layoffs.

I've scaffolded this report for you to make this process easier - as we talk about different sections of a report in class and read about how to create similar sections, you will practice by writing the equivalent section of your report.

The basic steps for this task are as follows:

- Identify the research question from the business question
- Identify data set(s) which are (1) publicly available (you don't have a budget to pay for private data) and (2) relevant to your task

    - (HW Week 6) Document your data sets in `draft-data-doc.qmd`

- Conduct a statistical analysis to support your answer to your research and business questions

    - Write a methods section for your business report corresponding to your statistical analysis
    - (HW Week 9) Draft of results section of business report with relevant graphics/visual aids in `draft-results.qmd`

- Write your report

    - (HW Week 10) Draft of Intro/Conclusion sections in `draft-intro-conclusions.qmd`

- – (HW Week 11) Draft of Executive summary section in `draft-exec-summary.qmd`

- Revise your report

  - – (HW Week 12 – not turned in) Revise your report

  - – (HW Week 13) - Rough draft of report due. Create one or more qmd files for your report (you can overwrite or delete intro.qmd and summary.qmd), include the names of each file (in order) in `_quarto.yml`. You should use references (edit references.bib and use pandoc citations). Make sure your report compiles and looks reasonable in both html and pdf.

  - – Develop a presentation to go along with your report (Week 13). Create slides for your report using quarto.

- Peer revise reports

  - – Peer revise reports

  - – (HW Week 14) - Make edits to your report from comments received from peer review

- Final report & presentation due

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

2 Summary

In summary, this book has no content whatsoever.

References

"IIF Databases : U.s. Bureau of Labor Statistics." n.d. Accessed February 17, 2026. https://www.bls.gov/iif/data.htm.

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.

A Draft: Data Documentation

A.0.1 Who collected the data:

The U.S. Bureau of Labor Statistics (BLS), with help from state labor agencies ("IIF Databases : U.s. Bureau of Labor Statistics" n.d.).

A.0.2 Why the data was collected:

To track nonfatal workplace injuries and illnesses nationally and help identify safety trends.

A.0.3 What the data is about:

Work related injuries, illnesses, and fatalities.Rates by industry, injury type, and severity.

A.0.4 When the data was collected:

Annually since 1972 and the most recent summary covers 2024.

A.0.5 Where the data was collected:

From all across the 50 states and a sample of ~200,000 employers.

A.0.6 How the data was collected:

Employers are legally required to log injuries and report to BLS.

A.0.7 Structure of the data:

Summary tables by industry and detailed case tables by event type and worker characteristics.

A.0.8 Formatting decisions:

Rates are per 100 FTE workers in summary tables and per 10,000 in detailed tables. Industry coded by NAICS, occupation by SOC. Cells under 15 cases are suppressed.

A.0.9 Data validation/quality control:

BLS runs consistency checks and publishes error estimates. Under reporting or failing to report by employers is a known limitation.

A.0.10 License:

U.S. government public domain.

A.0.11 Suggested analysis methods:

Compare injury and fall rates between manufacturing and office industries, use days lost for severity and rate ratios for risk.

A.0.12 Measurement instruments:

OSHA Forms 300/300A/301; BLS's OIICS coding system; NAICS and SOC classification systems.

A.1 References

B Draft: Results

C Draft: Intro/Conclusions

D Draft: Executive Summary