Project II

Due Date: March 25, 2025

Introduction

The goal of Project II is to develop your skills in applying intermediate mathematical and statistical methodologies, implementing data analysis and statistical computing algorithms, and conducting preliminary research. Additionally, you will enhance your ability to communicate data analysis findings effectively to both technical and non-technical audiences.

By the end of this project, you will be able to:

1. Apply intermediate mathematical and statistical methodologies to real-world problems.

2. Implement intermediate data analysis and statistical computing algorithms.

3. Use critical and analytical thinking to conduct preliminary research.

4. Communicate data analysis findings clearly and effectively to teammates and general audiences.

Instruction

Chapters 17 to 46 of "Data Science, Analytics, and AI for Business and the Real World" include 30 data science case studies and suggested solutions. Below is a list of the available case studies:

- Chapter 17 : Predicting the US 2020 Election
- Chapter 18 : Predicting Diabetes Cases
- Chapter 19 : Market Basket Analysis
- Chapter 20 : Predicting the World Cup Winner (Soccer/Football)
- Chapter 21 : Covid-19 Data Analysis and Flourish Bar Chart Race Visualization
- Chapter 22 : Analyzing Olympic Winners
- Chapter 23 : Is Home Advantage Real in Soccer and Basketball
- Chapter 24 : IPL Cricket Data Analysis
- Chapter 25 : Streaming Services (Netflix, Hulu, Disney Plus, and Amazon Prime)
- Chapter 26 : Micro Brewery and Pub Data Analysis
- Chapter 27 : Pizza Restaurant Data Analysis
- Chapter 28 : Supply Chain Data Analysis
- Chapter 29 : Indian Election Result Analysis
- Chapter 30 : Africa Economic Crisis Data Analysis
- Chapter 31 : Predicting Which Employees May Quit
- Chapter 32 : Figuring Out Which Customers May Leave
- Chapter 33 : Who to Target for Donations?
- Chapter 34 : Predicting Insurance Premiums
- Chapter 35 : Predicting Airbnb Prices
- Chapter 36 : Detecting Credit Card Fraud
- Chapter 37 : Analyzing Conversion Rates in Marketing Campaigns
- Chapter 38 : Predicting Advertising Engagement
- Chapter 39 : Product Sales Analysis

- Chapter 40 : Determining Your Most Valuable Customers
- Chapter 41 : Customer Clustering (K-Means, Hierarchical) - Train Passenger
- Chapter 42 : Build a Product Recommendation System
- Chapter 43 : Deep Learning Recommendation System
- Chapter 44 : Predicting Brent Oil Prices
- Chapter 45 : Detecting Sentiment in Tweets
- Chapter 46 : Spam or Ham Detection

The reference material is available in video format and can be accessed at:

https://julac-cuhk.primo.exlibrisgroup.com/permalink/852JULAC_CUHK/1dl2t2q/alma991040148765103407

The source code and datasets for each case study are available at:

https://github.com/PacktPublishing/Data-Science-Analytics-AI-for-Business-the-Real-World-

Tasks

Your tasks for Project II are as follows:

1. **Select and Replicate a Case Study:** Choose one case study from the provided list and replicate the analysis using the provided datasets and code.
2. **Download Relevant Data:** Obtain the datasets required for your chosen case study from the GitHub repository.
3. **Extend the Analysis:** Enhance the case study by applying additional statistical or computational methods (e.g., techniques learned in Project I or other STAT courses).
4. **Write a Report:** Prepare a detailed report in Jupyter Notebook format. The report should be tailored to a **general audience**. The report should include:
    - A clear problem statement.
    - A description of the methodology and tools used.
    - Code blocks that generate all results.
    - Visualizations and interpretations of key findings.
5. **Present Your Findings:** Deliver a 15-min presentation on either **March 25 (Week 12, Groups 1–5)** or **April 1 (Week 13, Groups 6–10)**. Your presentation should be tailored to a **general audience** and include:
    - The problem statement.
    - The approach used to address the problem.
    - Key findings and results.
    - A conclusion highlighting the strengths and limitations of your analysis.

Submission Requirements

1. **Project Materials:** Upload the following to Blackboard on or before **March 25, 2025**:
    - The dataset used for your analysis.
    - A presentation file (e.g., PowerPoint).
    - A Jupyter Notebook report.
2. **Programming Language:** Use Python for your analysis.

Assessment

Your performance will be evaluated based on the following criteria:

1. **Presentation (50%)**

    The presentation will be assessed on your ability to effectively communicate your findings to a general audience. Key aspects include:

    - **Clarity and Structure (25%):**

    ✓ Clear and logical organization of content.
    ✓ Effective use of visuals (e.g., charts, graphs) to support key points.
    ✓ Concise and well-articulated explanations.

    - **Engagement and Delivery (25%):**

    ✓ Ability to engage the audience and maintain their interest.
    ✓ Confidence and professionalism in delivery.
    ✓ Appropriate use of language and tone for a non-technical audience.

2. **Report (20%)**

    The Jupyter Notebook report will be evaluated based on the following:

    - **Quality of Analysis (10%):**

    ✓ Accuracy and rigor of the code and analysis.
    ✓ Appropriate use of statistical or computational methods.
    ✓ Clear and insightful interpretation of results.

    - **Organization and Clarity (10%):**

    ✓ Logical flow of ideas and analysis.
    ✓ Well-structured and easy-to-follow explanations.
    ✓ Professional presentation of visualizations and code.

3. **Extension of the Case Study Solution (20%)**

   This criterion evaluates your ability to creatively and meaningfully extend the original analysis:

   - **Creativity and Depth (10%):**

   ✓ Demonstration of innovative thinking in applying additional methods.
   ✓ Depth of exploration and justification for the chosen methods.

   - **Impact and Relevance (10%):**

   ✓ Clear explanation of how the proposed methods improve or complement the original analysis.
   ✓ Demonstrated understanding of the added value to the case study.

4. **Timeliness (10%)**

   Timeliness reflects your ability to meet deadlines and manage your time effectively:

   - **Checkpoint 1 (5%):**

   ✓ Submission by 23:59 on Feb 25:
   ✓ Selection of a case study.
   ✓ Downloading the solutions from GitHub.
   ✓ Allocation of work among team members.

   - **Checkpoint 2 (5%):**

   ✓ Submission by 23:59 on Mar 11:
   ✓ Proposal of additional statistical or computational methods.
   ✓ Explanation of why the proposed method(s) may improve the analysis.

Note: Final marks will be adjusted based on peer evaluation results.