WA3546-CAP Lab Guide

ADP Data Program Projects



© 2025 Ascendient, LLC

Revision 2.0.0 published on 2025-07-23.

No part of this book may be reproduced or used in any form or by any electronic, mechanical, or other means, currently available or developed in the future, including photocopying, recording, digital scanning, or sharing, or in any information storage or retrieval system, without permission in writing from the publisher.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

To obtain authorization for any such activities (e.g., reprint rights, translation rights), to customize this book, or for other sales inquiries, please contact:

Ascendient, LLC

1 California Street Suite 2900

San Francisco, CA 94111

https://www.ascendientlearning.com/

USA: 1-877-517-6540, email: getinfousa@ascendientlearning.com

Canada: 1-877-812-8887 toll free, email: getinfo@ascendientlearning.com

Table of Contents

1. Mi	id-Program Capstone Project: Employee Benefits Optimization for TechLance	4
1.1	1. Introduction	5
1.2	2. Scenario	5
	1.2.1. Benefits Optimization System	5
1.3	3. Resources	6
	1.3.1. Datasets	6
1.4	4. Capstone Instructions	6
	1.4.1. Phase 1: Data Foundation	6
	1.4.2. Phase 2: Exploratory Data Analysis	8
	1.4.3. Phase 3: Machine Learning	9
	1.4.4. Phase 4: Analytics Dashboard and Recommendations	
1 -	5. Conclusion	12

Mid-Program Capstone Project: Employee Benefits Optimization for TechLance

MODULE 1

1.1. Introduction

This capstone challenges you to develop a **comprehensive analytics solution** for TechLance Solutions, a fictional mid-sized company with 5,000 employees spending \$12M annually on employee benefits. The solution transforms raw benefits data into actionable insights to optimize the benefits portfolio, addressing underutilization, satisfaction gaps, and cost pressures. By integrating Data Science, Machine Learning, and Web Development, you will create a Python-based application to deliver insights and recommendations, executable via a single command-line call.

1.2. Scenario

TechLance Solutions, a mid-sized technology company, faces challenges in managing its \$12M benefits program:

- Cost Pressure: Benefits costs have risen 15% year-over-year.
- Low Engagement: Surveys indicate many employees underutilize available benefits.
- Satisfaction Gaps: Exit interviews highlight benefits as a retention concern.
- Resource Allocation: Lack of clear data on benefits' return on investment (ROI).
- Demographic Shifts: A multi-generational workforce with diverse needs.

The HR leadership team seeks a data-driven solution to optimize benefits, improve employee satisfaction, and enhance retention while controlling costs.

1.2.1. Benefits Optimization System

A Python-based web application (e.g., Flask, Streamlit or Gradio) will integrate data processing, analytics, machine learning, and a dashboard to provide actionable insights. The application will run from a single command (e.g., python app.py) and

support caching for faster load times. It must handle new data entries with the same schema seamlessly.

1.3. Resources

1.3.1. Datasets

The following datasets are provided in the data directory: * usage_data.csv: Contains EmployeeID, BenefitID, UsageFrequency, LastUsedDate.

- employee_data.csv: Includes EmployeeID, Age, Gender, Department,
 Tenure.
- benefits_data.csv: Details BenefitID, BenefitType, BenefitSubType, BenefitCost.
- feedback_data.csv : Provides EmployeeID , BenefitID , SatisfactionScore , Comments .

1.4. Capstone Instructions

This capstone is open-ended, encouraging creative solutions using Data Science, Machine Learning, and Generative AI. The following sections outline the key phases and requirements.

1.4.1. Phase 1: Data Foundation

Build a robust data pipeline to ensure high-quality, integrated data for analysis.

Note

It is recommended to create a python notebook .ipynb file to organize your thoughts, clean the data, and prepare the data foundation and EDA exploration of the results.

1.4.1.1. Task 1.1: Data Quality Assessment

- Load datasets (usage_data.csv), employee_data.csv, benefits_data.csv, feedback_data.csv) using pandas.
- 2. Profile data to identify issues (e.g., missing values, duplicates, outliers like high UsageFrequency).
- 3. Document findings and propose solutions (e.g., impute missing SatisfactionScore, remove duplicates).
- 4. Define validation rules (e.g., SatisfactionScore in 1–5, LastUsedDate in 2023–2025).

Deliverable: 1–2 paragraph summary in a python notebook (not required in the final app).

1.4.1.2. Task 1.2: Data Integration and Preparation

- 1. Merge datasets using <a>EmployeeID and <a>BenefitID to create a unified dataset.
- 2. Handle missing values (e.g., impute UsageFrequency with median, exclude incomplete Comments).
- 3. Standardize data types (e.g., LastUsedDate to datetime, encode Gender / Department).
- 4. Create derived fields:
 - Age groups (Gen Z: 18–25, Millennial: 26–41, Gen X: 42–57, Boomer: 58+).
 - Tenure categories (<5 years, 5–15 years, 15+ years).
 - Benefit subcategory flags (e.g., "401k Basic Matching").
- 5. Validate the dataset (e.g., no negative UsageFrequency).

Deliverable: Clean, integrated dataset (CSV) and notebook documenting preprocessing.

1.4.2. Phase 2: Exploratory Data Analysis

Analyze data to uncover utilization patterns, demographic trends, and cost efficiency.

1.4.2.1. Task 2.1: Utilization Pattern Analysis

- 1. Analyze usage_data.csv for most/least utilized benefits by UsageFrequency.
- 2. Examine temporal trends in LastUsedDate (e.g., monthly usage for 2023–2024).
- 3. Categorize employees by engagement (high: UsageFrequency ≥ 7, medium: 3–6, low: 0–2).
- 4. Analyze utilization by BenefitSubType (e.g., "401k Basic Matching" vs. "401k High Contribution").
- 5. Create visualizations (e.g., bar charts, line plots) using Plotly.

Deliverable: Utilization Analysis Dashboard (5–7 visualizations with analysis of trends and patterns).

1.4.2.2. Task 2.2: Demographic Insights Discovery

- Analyze benefit usage by demographics (Age , Gender , Department , Tenure).
- 2. Identify preferences by BenefitSubType (e.g., IT's preference for "Technology Stipend").
- 3. Create visualizations (e.g., stacked bar charts, heatmaps).
- 4. Document differences (e.g., Gen Z's preference for "Gym Membership").

Deliverable: Demographic Analysis Report with visualizations.

1.4.2.3. Task 2.3: Cost Efficiency and Subcategory Analysis

- Calculate cost-per-usage (BenefitCost ÷ UsageFrequency) by BenefitID / BenefitSubType
- 2. Develop an ROI score (normalize cost-per-usage and SatisfactionScore).
- 3. Identify underutilized high-cost subcategories (e.g., "Gym Membership: Tier 3 Partners").
- 4. Rank benefits by cost efficiency, flagging candidates for elimination.
- 5. Create quadrant plots (cost vs. usage, cost vs. satisfaction).

Deliverable: Cost Efficiency Analysis Report with recommendations.

1.4.3. Phase 3: Machine Learning

Apply machine learning to analyze satisfaction, segment employees, and recommend benefits.

1.4.3.1. Task 3.1: Employee Satisfaction Analysis

- 1. Analyze SatisfactionScore distributions by BenefitID, BenefitSubType, and demographics.
- 2. Compute correlations between UsageFrequency and SatisfactionScore.
- 3. Perform predictive sentiment analysis on Comments (positive, neutral, negative).
- 4. Identify satisfaction drivers (e.g., regression on SatisfactionScore).
- 5. Create scorecards by BenefitType / BenefitSubType with sentiment insights.

Deliverable: Employee Satisfaction Insights Report.

1.4.3.2. Task 3.2: Employee Segmentation

1. Create usage vectors (total UsageFrequency per BenefitSubType).

- 2. Generate temporal profiles (e.g., monthly usage via PCA or summary stats).
- 3. Apply clustering (e.g., K-Means with silhouette score, Gaussian Mixture Models).
- 4. Validate clusters (silhouette score, Davies-Bouldin index).
- 5. Profile clusters by usage patterns and demographics (e.g., "wellness enthusiasts").

Deliverable: Segmentation Analysis Report with visualizations.

1.4.3.3. Task 3.3: Recommender System

- 1. Build a user-item matrix (EmployeeID vs. BenefitID / BenefitSubType , values
 = UsageFrequency).
- 2. Use collaborative filtering (e.g., k-NN, SVD) or content-based filtering (e.g., cosine similarity on metadata).
- 3. Evaluate offline (Precision@K, Recall@K, MAP) and propose an A/B test for online evaluation.
- 4. Suggest benefits based on peer usage or metadata (e.g., department, tenure).

Deliverable: Recommender System Notebook with evaluation metrics and sample recommendations.

1.4.4. Phase 4: Analytics Dashboard and Recommendations

Integrate insights into an interactive dashboard and provide strategic recommendations.

1.4.4.1. Task 4.1: Business Intelligence Dashboard

1. Build a dashboard (Streamlit/Gradio) with metrics (utilization, satisfaction, segments, ROI).

- 2. Include interactive filters (e.g., Department , Age , BenefitSubType).
- 3. Incorporate Generative AI insights (e.g., feedback summaries) in tooltips/annotations.
- 4. Ensure mobile-friendly design (e.g., using Plotly Dash).
- 5. Export as interactive HTML or PDF.

○ Tip

Implement caching (e.g., for processed datasets) to improve load times for repeated runs.

Deliverable: Interactive BI Dashboard (standalone deployed application that runs from command line).

1.4.4.2. Task 4.2: Strategic Recommendations Report

- 1. Synthesize findings, emphasizing subcategory analysis (e.g., cut "Gym Membership: Tier 3 Partners").
- 2. Prioritize recommendations by impact (e.g., cost savings) and feasibility.
- 3. Provide a 6-month roadmap (e.g., pilot new benefits, phase out low-ROI ones).
- 4. Estimate cost savings (e.g., 10–20% of \$12M) and ROI improvements.
- 5. Address stakeholder needs (HR: retention, Finance: savings, Leadership: strategy).
- 6. Include risk assessment (e.g., employee pushback).

Deliverable: Strategic Recommendations Report based on synthesized insights from the 4 phases.

1.5. Conclusion

In this capstone, you will design and implement a **Benefits Optimization System** to help TechLance Solutions optimize its \$12M benefits portfolio. By leveraging Data Science, Machine Learning, and Generative AI, you will deliver actionable insights to improve utilization, satisfaction, and cost efficiency. The final application should integrate all components into a user-friendly, command-line-executable tool for HR and leadership.