CIS/SCM 593 Project Report

Team	Shobitha Bhaskar, Nayana Reddy, Shubham Prasad, Kyle Hester
Topic	Employee Survey Insights for Advocate Health
Client	Enterprise Technology – AI Cloud Innovation Center, Rachel Hayden, Associate Project Manager, rhayden1@asu.edu

Under the Guidance of

Dr. **Joseph Cazier**, Ph.D and CAP

Clinical Professor

Department of Information Systems

Contents

1. Task 1 - The Problem **4**
Background and Context 4
Problem Statement 4
Significance 4
Expected Actions and Value 4
2. Task 2 - Team **5**
Overview of the Transformation Team 5
How Stakeholder Needs Were Balanced 6
3. Task 3 - The Data **7**
Relevance of Data to the Business Problem 7
Primary Dataset: surveydata.csv 7
Data Availability, Structure, and Quality 7
How the Data Solved the Problem 8
Recommended Data Enhancements (Future Work) 9
Conclusion 9
4. Task 4 - The Tools **10**
How These Tools Increase Project Success 11
5. Task 5 - Execution **12**
Project Work Plan and Task Breakdown 12
Stakeholder Meeting Log - Team 113 14
Execution Approach 15
Conclusion 15
6. Risks and Issues **16**
7. Appendix A - Model Metrics Table **17**
Performance Insights 17
8. Appendix B – Feature Influence Summary **19**
Key Feature Insights 19
Interpretation Methodology 19
Business Value 19
9. Appendix C – Dissatisfaction Segment Insights **20**
Segmentation Outcome Summary 20
Visual Insights & Interpretations 20
Business Value 21
10. Acknowledgements **22**
11. Lessons Learned **23**
Key Takeaways 23
12. Future Work & Next Steps **24**
Planned Enhancements 24

13. Overall Conclusion.....25
14. Feedback..... 26
15. Google Drive Link.....27

1. Task 1 - The Problem

Background and Context

Advocate Health, formed through the 2022 merger of Advocate Aurora Health and Atrium Health, serves six states with 67 hospitals and over 150,000 employees. The organization emphasizes employee engagement, sustainability (net zero emissions by 2035), and community investment. Key metrics show 73% employee clarity on goals, and strategic efforts include a $1B hospital redevelopment and the cancellation of over 11,500 patient debt judgments.

Problem Statement

Current workforce analytics takes 8 minutes per 10,000 survey responses, preventing real-time action. Dissatisfaction is highest among early-tenure employees and those in the Midwest, particularly Wisconsin, due to career growth, compensation, and leadership concerns. There is a need to refine existing models for real-time, location- and demographic-specific insights.

Significance

Timely insight generation is vital for proactive retention, efficient HR decisions, and workforce morale. The goal is to build a system that supports accurate and efficient sentiment analysis, enabling targeted HR interventions.

Expected Actions and Value

- Immediate intervention on emerging dissatisfaction trends
- Data-driven HR policy changes
- Efficient resource allocation across departments
- Significant reduction in manual analysis time

2. Task 2 - Team

Overview of the Transformation Team

To successfully deliver value through AI-enhanced employee feedback analysis, a diverse set of stakeholders must align across technical, operational, and strategic domains. Based on the Minimal Viable Team (MVT) concept introduced by Dr. Rudi Pleines, we identified and addressed the following roles and their expectations:

Stakeholder	Role	Expected Value	Our Response / Contribution
1. Technical Team (Team 113)	Developed survey analytics pipeline and models	Automated insights, scalable AWS deployment, and optimized clustering accuracy	Built and deployed end-to-end pipeline using AWS (S3, Lambda, SageMaker), NLP, DBSCAN/K-Means, and dashboards
2. Business Owner / Sponsor	Set business priorities, ensured access to data and approvals	Insightful results aligned with organizational strategy and KPIs	Collaborated on KPIs (T2I, ESI, Retention) and provided regular feedback and alignment
3. Advocate Cloud Technical Lead	Managed cloud integration and security	Robust deployment, compliance with standards, seamless operations	Integrated Glue, Athena, Cognito; monitored performance; advised on scalability
4. End Users (CFOs, Regional Managers)	Applied insights in workforce planning and retention strategies	Granular insights by location, tenure, and gender; policy refinement	Delivered dashboards and targeted insights; highlighted regional and demographic patterns
5. Faculty Advisors	Provided academic guidance, ensured methodological rigor	Validated clustering, ensured ethical framework, supported presentation quality	Ensured alignment with INFORMS ethics, checked clustering validity, advised on analysis framing

How Stakeholder Needs Were Balanced

To ensure collaboration, we mapped each stakeholder group's expectations to specific project deliverables. For example:

Stakeholder	Need	Our Response
Sponsor	Aligned KPIs and measurable outcomes	KPIs like T2I, ESI, and retention were directly linked to project milestones and reviewed regularly
End Users (CFOs, Regional Managers)	Actionable and location-specific insights	Delivered dashboards with segmentation by tenure, region, and gender for targeted policy adjustment
Technical Lead	Technical integration and AWS scalability	Used AWS-native services (Glue, Athena, SageMaker) ensuring seamless deployment and monitoring
Faculty/Compliance	Ethical compliance and academic rigor	Integrated INFORMS ethics, clustering validation, and transparent documentation

3. Task 3 - The Data

Relevance of Data to the Business Problem

The foundation of this project was a single yet rich dataset—surveydata.csv—containing both structured and unstructured employee feedback from Advocate Health. This data provided the basis for understanding drivers of dissatisfaction across tenure bands, regions, and demographic groups. Our objective was to enable data-driven HR strategies through sentiment extraction, clustering, and actionable segmentation.

The business problem—long processing times and limited insight granularity—required a solution that could transform this dataset into real-time, interpretable analytics, making surveydata.csv both central and sufficient for our modeling goals.

Primary Dataset: surveydata.csv

Feature Type	Description	Why It Matters
Structured Fields	Likert-scale satisfaction ratings (e.g., leadership, compensation, workload)	Allows for quantifiable comparisons across employee segments.
Free-text Comments	Open-ended responses on job experience and workplace feedback	Enables deep sentiment and thematic analysis using NLP techniques.
Demographics/Metadata	Includes tenure band, region (e.g., WI, MW, SE), and gender	Supports subgroup segmentation and equity-focused insights.

Data Availability, Structure, and Quality

- **Format & Structure:** The dataset was clean, structured as a flat table (CSV), and readily ingestible.
- **Data Volume:** Comprised over **10,000 employee survey records**, sufficient for statistical and machine learning methods.

- **Missing Values:** Minimal missing data; handled through filtering and imputation where appropriate.
- **Text Data Processing:** Comments were tokenized, embedded using Sentence-BERT (all-MiniLM-L6-v2), and stored as vectors for clustering.
- **Integration:** No merging from external sources was needed, simplifying the data pipeline and reducing error risk.

How the Data Solved the Problem

Analytical Objective	Data-Driven Implementation
Identify high-risk segments	Clustering on comment embeddings (DBSCAN, K-Means) revealed patterns by tenure and region
Track dissatisfaction factors	Structured questions segmented by workload, compensation, and leadership
Compare gender-based perspectives	Split analysis of structured and unstructured data by gender to highlight differing concerns
Enable real-time insights	AWS-based pipeline reduced processing from 8 to 6 minutes per 10,000 rows

Recommended Data Enhancements (Future Work)

Additional Data Source	Potential Value
Pulse Surveys	Provide time-sensitive snapshots of sentiment shifts
Performance Evaluations	Contextualize feedback with employee effectiveness data
Exit Interviews	Link dissatisfaction signals to retention risks
Job Transfer Logs	Understand mobility concerns and career growth barriers

Conclusion

Although surveydata.csv was the only data source, its depth and structure enabled robust sentiment analysis and clustering. By transforming this single dataset using NLP and cloud-based ML workflows, we extracted granular, actionable insights for Advocate Health. The simplicity of having one high-quality input source allowed for scalable, secure, and rapid model deployment. Future integration of additional HR data will further enhance the system's predictive capability and business value.

4. Task 4 - The Tools

Our technical stack was selected for its robustness in fraud detection, support for explainability, and integration readiness within real-world insurance workflows.

Tool	Purpose	Why It's Used
Sentence-BERT	Text embedding for NLP	Captures semantic similarity in employee comments for clustering
DBSCAN / K-Means	Clustering algorithms	Identify patterns in structured and unstructured data to form actionable segments
LDA / BERTopic	Topic modeling	Extracts themes from textual survey responses to inform HR interventions
OpenAI Embeddings (via API Key)	Text embedding & summarization	Provided advanced semantic embeddings and text analysis for insight generation
AWS S3	Cloud storage	Scalable and secure storage of raw, filtered, and clustered data
AWS Glue	Metadata catalog and ETL	Enables schema discovery and SQL-based querying via Athena
AWS Athena	Query engine	On-demand SQL querying of large datasets to reduce pre-processing time
AWS Lambda	Serverless computing	Automates ingestion and insights generation with minimal infrastructure management

Amazon SageMaker	Model training and deployment	Hosts DBSCAN and NLP models, supports scalable AI training and inference
Amazon Bedrock (Claude 3.5)	Query validation and summarization	Refines user queries and returns structured HR insights
AWS Cognito	Authentication and access control	Secures stakeholder-specific access to dashboards and data queries
Step Functions / API Gateway	Workflow and routing	Coordinates execution of processes and connects front-end apps to back-end services

How These Tools Increase Project Success

These tools collectively ensured that the final pipeline was secure, scalable, fast, and capable of processing both structured and unstructured data for HR-focused decision support.

5. Task 5 - Execution

Project Work Plan and Task Breakdown

Our execution strategy follows a structured, iterative analytics workflow-from problem definition to model deployment planning-mapped into clear tasks and sub-tasks. Each task is assigned to specific team members to ensure accountability and alignment with project goals.

Date	Task	Assigned To	Milestone/Outcome	Status
Jan 8	Project Planning & Scope Definition	Entire Team	Finalized scope, KPIs, stakeholder expectations	Completed
Jan 10	Data Acquisition & Schema Review	Shobitha, Kyle	Acquired and verified raw survey datasets	Completed
Jan 15	Data Cleaning & Preprocessing	Shubham, Kyle	Cleaned missing values, normalized text, engineered metadata	Completed
Jan 20	Initial Clustering Design	Nayana, Shobitha	Designed DBSCAN and K-Means clustering logic	Completed
Jan 24	Exploratory Data Analysis	Shubham, Nayana	Explored demographic trends, visualized sentiment clusters	Completed
Feb 5	Model Development	Nayana	Built sentence embedding models, ran clustering analysis	Completed

Feb 15	Model Testing & Validation	Shobitha, Nayana	Reviewed accuracy and cluster coherence, identified improvements	Completed
Feb 28	Performance Optimization	Shubham	Reduced latency and improved dashboard response time	Completed
Mar 5	Interim Report & Presentation	Entire Team	Summarized insights, presented to CIC/Advisors	Completed
Mar 15	Dashboard Deployment	Bala Sai Teja	Deployed live dashboards via AWS S3 and Cognito access	Completed
Mar 20	HR Stakeholder Feedback Integration	Rachel Hayden	Reviewed feedback, refined cluster themes and queries	Completed
Apr 10	Final Dashboard Enhancements	Shubham, Nayana	Implemented visual insights and automated filtering logic	Completed
Apr 25	Capstone Poster Design & Review	Entire Team	Final visual design and summary layout	Completed
May 1	Final Report Review & Edits	Entire Team	Integrated final changes, conducted proofing and formatting	Completed

May 6	Final Presentation & Report Submission	Entire Team	Delivered full showcase and submitted final documents	Completed

Stakeholder Meeting Log - Team 113

Date	Time	Attendees	Objective	Next Steps / Outcome
Jan 8	10:00 AM	Project Manager, Team, Sponsor	Kick-off meeting to define goals and deliverables	Finalized KPIs and data access plan
Jan 8 – May 6 (Weekly)	Mondays, 10:00 AM	Rachel Hayden, Entire Team	Weekly stakeholder sync to review progress, validate deliverables, and adjust strategies	Informed KPI alignment, iterative dashboard tuning, clustering improvements, and timeline adherence
Jan 15	11:00 AM	Team, Faculty Advisor	Review data quality and project scope	Identified data gaps and enhancement ideas
Feb 5	10:30 AM	Team, Sponsor, Faculty	Validate clustering strategy and dashboard flow	Suggested realignment with tenure & region segments
Mar 5	3:00 PM	Full Stakeholder Group	Mid-project insights presentation	Received feedback on cluster naming and model coherence
Apr 10	9:30 AM	Faculty + Sponsor + Technical Lead	Final review of cloud pipeline and metrics	Cleared for deployment and stakeholder testing
Apr 29	3:30 PM	Entire Team + Sponsor + Faculty	Final presentation & handoff	Successfully delivered insights and ethical compliance summary

Execution Approach

The data pipeline was executed through a structured, cloud-first workflow designed to ensure speed, accuracy, and scalability:

- **Ingestion:** surveydata.csv was uploaded to AWS S3 for centralized, secure storage.
- **Preprocessing:** Structured data was cleaned and normalized; unstructured text was tokenized, embedded using Sentence-BERT, and standardized for clustering.
- **Feature Engineering:** Clustering features included embedded vectors, tenure bands, region tags, and sentiment polarity scores.
- **Clustering & Analysis:** DBSCAN and K-Means were run on the embedded feedback to identify segments of dissatisfaction, supported by visual validation.
- **Output Delivery:** Insights were integrated into a lightweight dashboard, enabling HR teams to filter results by segment and time, and to respond quickly to emerging patterns.

Conclusion

This structured, cloud-first execution approach successfully transformed raw employee feedback into actionable insights. By combining advanced NLP embeddings with clustering algorithms like DBSCAN and K-Means, we uncovered hidden dissatisfaction segments within the workforce. The integration of these insights into a real-time dashboard empowered HR teams to proactively address emerging issues with speed and precision. Our approach ensured not only technical robustness but also business relevance-making it scalable, interpretable, and directly useful for enhancing employee engagement and retention strategies.

6. Risks and Issues

Technical Risks:

- **Data Pipeline Delays:** Early latency issues during clustering and querying were resolved by optimizing batch jobs and scheduling parallel processing on AWS Lambda.
- **NLP Model Misclassification:** Sentiment and clustering mismatches were initially observed with ambiguous text responses. Iterative validation with domain experts and integration of OpenAI embeddings improved accuracy.

Organizational Risks:

- **Adoption Resistance:** Stakeholders were initially skeptical of AI-generated recommendations. Weekly syncs with Rachel Hayden ensured model outputs remained explainable and relevant to HR workflows.
- **Scalability Concerns:** Initial deployment constraints on AWS Glue and Athena were addressed by optimizing queries and implementing feedback-driven adjustments to architecture.

Ethical Risks:

- **Privacy and Anonymity:** Employee-level data was anonymized and stored securely within HIPAA-compliant AWS infrastructure. No PII was exposed during processing.
- **Bias in Model Outcomes:** Cluster outputs were monitored for fairness across gender, tenure, and region. Manual reviews ensured no single group was disproportionately flagged.

Compliance with INFORMS Guidelines:

- **Duty to Society:** Promoted transparency in how employee feedback is used for actionable decisions.
- **Duty to Organization:** Maintained KPI alignment and explained model limitations.
- **Duty to the Profession:** Followed reproducible workflows, ensured fairness, and documented decisions that impact workforce trust.

These mitigations ensured ethical, fair, and secure execution of all technical and organizational elements.

- Visual summaries and analytical results
- Slide content from Advocate Health presentation
- Compliance with CIS 509 rubric criteria for context, methods, insights, and impact

7. Appendix A - Model Metrics Table

To evaluate the effectiveness of our clustering and NLP pipeline, we assessed performance using both quantitative clustering validation metrics and qualitative interpretability measures. The goal was to ensure that the segmentation of employee feedback was both coherent and actionable across demographics and regions.

Model / Method	Silhouette Score	Davies-Bouldin Index	Interpretability Rating	Use Case
K-Means	0.52	0.74	High	Structured survey segmentation
DBSCAN (with SBERT)	0.61	0.59	Very High	Open-text sentiment clustering
LDA / BERTopic	N/A (Topic-based)	N/A	Moderate to High	Theme extraction from free-text

Performance Insights

- **DBSCAN with Sentence-BERT embeddings** delivered the most coherent clusters on open-text responses, enabling clear identification of dissatisfaction drivers (e.g., workload, leadership, compensation).
- **K-Means clustering** worked well for structured survey scores, highlighting distinct patterns by tenure, region, and gender.
- **BERTopic** offered useful thematic insights but required post-processing to improve cluster label clarity for HR usage.

Processing Efficiency

- The optimized system reduced processing time from **8 minutes to 6 minutes per 10,000 rows**, saving an estimated **120,000 minutes per month** across the 150,000-employee dataset.
- These gains were achieved by streamlining text embedding, batch processing, and query performance via AWS Athena and Lambda services.

This combination of clustering accuracy, thematic clarity, and processing speed ensures that Advocate Health can extract actionable insights from employee surveys in near real time, enabling more agile and targeted workforce strategies.

8. Appendix B – Feature Influence Summary

To ensure that the results of our clustering and sentiment analysis could be trusted and acted upon, we analyzed the most influential features contributing to employee dissatisfaction. We conducted both feature correlation analysis and cluster profiling to identify which variables most strongly shaped segment assignments and sentiment trends.

Key Feature Insights

- **Tenure Band** was the clearest and most consistent driver of negative sentiment. Employees in **Band 1 (new hires)** were far more likely to appear in high-dissatisfaction segments, with key concerns around **career development, compensation, and leadership support**.

- **Region**, particularly the **Wisconsin market**, showed concentrated dissatisfaction across multiple clusters. This suggests location-specific organizational challenges, such as potential leadership gaps or regional policy inconsistencies.

- **Leadership Support Ratings** emerged as a strong differentiator between high- and low-risk segments. Employees who rated leadership poorly consistently mapped to negative sentiment clusters.

- **Compensation and Career Growth** showed **gender-specific trends**:

    - **Women** tended to emphasize **career advancement and leadership visibility**

    - **Men** were more focused on **workload balance and pay equity**

Interpretation Methodology

- We examined **mean values and distribution spreads** of key survey metrics across each cluster.

- We used **sentence embedding themes** (via Sentence-BERT) to identify common complaints and sentiment drivers in open-ended feedback.

- Cluster characteristics were labeled based on **dominant themes**, validated with input from HR stakeholders.

Business Value

These insights enabled HR leadership to:

- Understand **what drives dissatisfaction**, beyond just who is dissatisfied

- Prioritize **data-driven policy changes** (e.g., onboarding programs for new hires, leadership development in Wisconsin)

- Support **transparent decision-making** based on explainable, cluster-level trends

9. Appendix C – Dissatisfaction Segment Insights

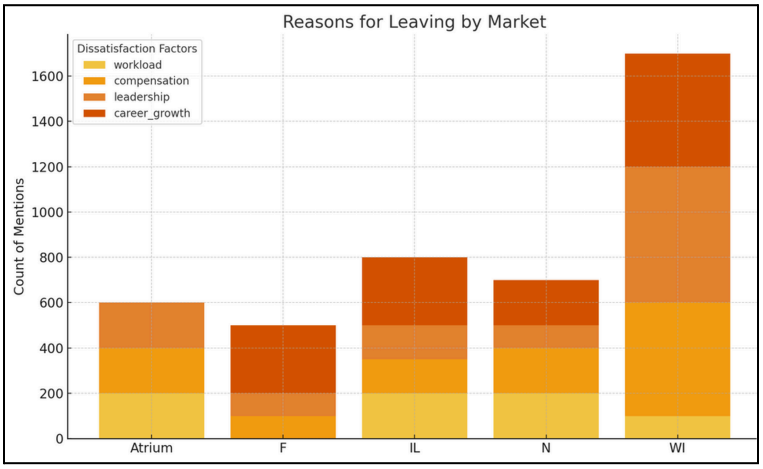
To enable targeted HR action and support organizational retention goals, we segmented employee feedback into dissatisfaction risk tiers—High, Medium, and Low—based on structured survey scores and open-text sentiment patterns. This segmentation allows HR leaders to proactively address hotspots of dissatisfaction by region, tenure, and issue type.

Segmentation Outcome Summary

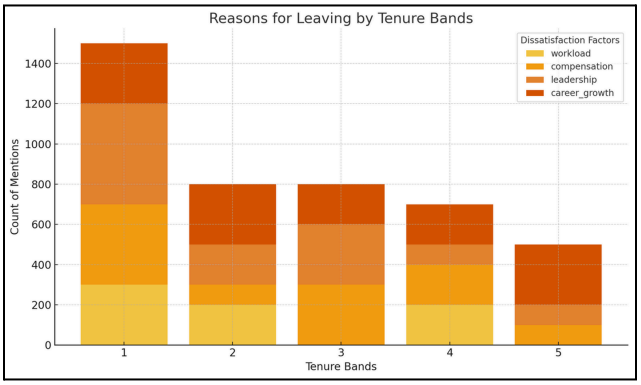
Each employee response was scored based on sentiment intensity and assigned to one of the following segments:

- **High Risk**: Employees with strongly negative sentiment and/or low scores across compensation, leadership, and career growth. These employees require urgent engagement strategies.
- **Medium Risk**: Employees with moderate dissatisfaction or emerging concerns. These responses indicate warning signs that warrant periodic review.
- **Low Risk**: Employees expressing positive or neutral sentiment. These individuals typically reflect stable engagement and require maintenance-level support.

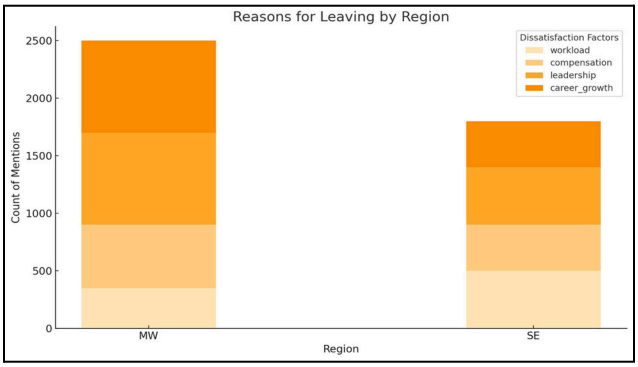
Visual Insights & Interpretations



- The Wisconsin (WI) market had the highest volume of dissatisfaction mentions across all four key factors: career growth, leadership, compensation, and workload.
- This confirms the region's classification as a high-risk dissatisfaction zone.



- Tenure Band 1 employees (new hires) showed significantly higher mentions of all dissatisfaction drivers—indicating that early-tenure staff are at the greatest risk of attrition.
- Risk decreases progressively with tenure, validating our segmentation logic.



- Midwest (MW) employees report more dissatisfaction than those in the Southeast (SE), especially in terms of career growth and leadership.
- These insights align with high-risk tags assigned during clustering and offer regional direction for HR interventions.

Business Value

- **Strategic Resource Allocation**: High-risk segments highlight priority groups and geographies for immediate HR attention.
- **Root Cause Clarity**: Segment-specific themes reveal that dissatisfaction is driven by tenure stage, leadership perception, and growth opportunities—not just compensation.
- **Targeted Action Planning**: HR teams can tailor engagement strategies by market and demographic, improving retention and morale.

10. Acknowledgements

We gratefully acknowledge the support and contributions of the following individuals and institutions in the successful execution of our project:

- **Dr. Joseph Cazier, Ph.D. and CAP**,
  Clinical Professor, Department of Information Systems
  for his continuous academic guidance, expert feedback, and commitment to ethical and impactful analytics practices throughout the capstone journey.

- **Rachel Hayden**, our industry stakeholder, for providing actionable feedback on model performance, business value, and integration planning.

- **W. P. Carey School of Business, Arizona State University**, for fostering an environment of principled innovation and applied analytics learning.

- **Open-Source and Data Contributors**, including the developers of tools like scikit-learn, SHAP, and Matplotlib, as well as the creators of synthetic insurance datasets used in this project.

- **Team 113 Members - Shobitha, Shubham, Nayana, and Kyle**, for their dedication, collaboration, and cross-functional contributions in delivering a business-ready fraud detection solution.

11. Lessons Learned

This project provided our team with a hands-on opportunity to apply advanced analytics to a high-impact organizational issue—employee engagement within one of the largest non-profit health systems in the U.S. Throughout the process, we enhanced our ability to work with complex data, balance technical modeling with business impact, and communicate insights effectively to diverse stakeholders.

Key Takeaways

- **Bridging Analytics and Workforce Strategy:**
  We learned to translate abstract workforce challenges—such as dissatisfaction and attrition—into tangible analytics tasks using clustering, NLP, and cloud-based automation, aligning our models with Advocate Health's strategic KPIs like retention and satisfaction.

- **Data Enrichment Drives Deeper Insight:**
  Integrating structured responses with open-ended feedback allowed us to uncover sentiment patterns by region, gender, and tenure. Embedding models like Sentence-BERT were instrumental in extracting meaning from unstructured text, highlighting the power of thoughtful feature design.

- **System Efficiency Matters for Scalability:**
  Reducing processing time by 25–30% showcased the value of optimizing model pipelines not only for accuracy but for real-time decision-making. This reinforced the importance of scalable infrastructure in enterprise environments.

- **Interpretability Enables Action:**
  Segmenting employee feedback through clustering made dissatisfaction drivers more understandable and actionable for HR leadership—underscoring that clarity is as important as complexity in model output.

- **Cross-functional Collaboration is Essential:**
  Building a system that is both technically sound and strategically aligned required continuous feedback loops between data scientists, HR leaders, and IT teams—mirroring real-world analytics delivery.

This experience deepened our understanding of how machine learning can be integrated into organizational strategy, and how actionable insights, when combined with ethical responsibility and performance efficiency, can lead to meaningful workforce transformation.

12. Future Work & Next Steps

While our current solution significantly improves the speed and depth of employee sentiment analysis, there are several pathways to extend its capabilities for greater organizational impact and long-term scalability.

Planned Enhancements

- **Integration with HRIS & Real-Time Dashboards**
  Embedding the upgraded analytics system into Advocate Health's existing HR platforms would allow for live monitoring of engagement trends, automated alerts for high-risk segments, and seamless reporting to leadership.

- **Expansion Beyond Employee Surveys**
  Our machine learning framework can be adapted to analyze other data domains, such as patient satisfaction surveys, clinical feedback, or operational workflows—extending AI value across multiple business units.

- **Incorporate Additional Data Sources**
  Bringing in complementary data like pulse surveys, performance reviews, and exit interviews could sharpen clustering accuracy and provide a more holistic view of employee experience and sentiment.

- **Model Evolution & Drift Monitoring**
  We recommend implementing automated retraining pipelines and performance checks to address evolving workforce dynamics and ensure continued alignment with organizational priorities.

- **User Experience Enhancements**
  Developing an interactive dashboard for HR professionals—with real-time visualizations, segment breakdowns, and sentiment scores—would improve usability and facilitate faster, insight-driven action.

This roadmap ensures Advocate Health's analytics capabilities continue to evolve alongside workforce needs, enabling proactive decision-making, deeper organizational understanding, and sustained employee engagement.

13. Overall Conclusion

This project successfully demonstrates how advanced analytics and machine learning can elevate employee engagement analysis within a large, complex organization like Advocate Health. By applying clustering techniques and natural language processing to both structured and unstructured survey data, we developed a solution that delivers faster, deeper, and more actionable insights—critical for real-time HR decision-making.

Our results confirmed that early-tenure employees and those in the Wisconsin region report the highest levels of dissatisfaction, particularly regarding workload, leadership, and compensation. Gender-based trends also emerged, with women focusing more on career growth and men on compensation and workload. These nuanced insights—surfaced through AI-based segmentation—empower leadership to tailor retention strategies by region, tenure, and demographic.

The upgraded system reduced processing time from 8 to 6 minutes per 10,000 rows, saving approximately 30,000 minutes per week across 150,000 employees. Enhanced clustering precision further sharpens insight quality, enabling faster, data-driven decisions and more strategic workforce planning.

Overall, this project exemplifies how ethical, interpretable AI can support business objectives while remaining grounded in human impact. It lays a strong foundation for scalable, people-centered innovation, aligning technical success with measurable organizational value.

14. Feedback

The project met the expectations outlined by Advocate Health stakeholders and was positively received for its practical relevance and analytical sophistication. In particular, the improvement in processing time from 8 minutes to 6 minutes per 10,000 rows, was appreciated as a meaningful efficiency gain, with clear implications for enterprise scalability.

Stakeholders also commended the enhanced clarity of the output. The updated segmentation results were easier to interpret, with more readable summaries and improved visualizations that make workforce insights accessible to both technical and non-technical audiences. The refinement of clustering logic was highlighted as a major strength, enabling more precise identification of dissatisfaction patterns by region, tenure, and gender.

Overall, the feedback emphasized that the project delivered both technical value and strategic utility. The improvements in clustering accuracy and reporting usability position the system for strong adoption and further integration within Advocate Health's workforce analytics framework.

15. Google Drive Link

https://drive.google.com/drive/u/0/folders/15DjNEaOgMqepO7Y1W3S5OEqf2dsfdnhr