AMS: Database Migration and Application

The Data Mine Corporate Partners Symposium 2025

Students: Richard Lin, Dhara Inani, Caroline Finch, Kaden Drew Furigay, Evan Zhang, Victoria Jung

AMS | AMERICAN MATHEMATICAL SOCIETY

PURDUE UNIVERSITY. The Data Mine

Introduction

Project Goal:

Migrate historical survey data from FileMakerPro (discontinued by AMS) into a modern database. This will allow the AMS to import the data into their relational database and support the creation of a public-facing dashboard for interacting with the data and exploring trends.

- Exporting data from FileMakerPro to Excel is inefficient, requires repeated manual effort
- FileMakerPro does not support direct integration with modern analytics tools
- PostgreSQL allows for automated data flow and real-time updates

PostgreSQL and File Migration

- Created PostgreSQL database
- Exported FileMaker Pro files to a .csv format
- Imported .csv files into Postgres database
- Provided database access to other teams
- Merged EENDR and DGData 2010-13 into a single table, to assist visualization team with making visualizations
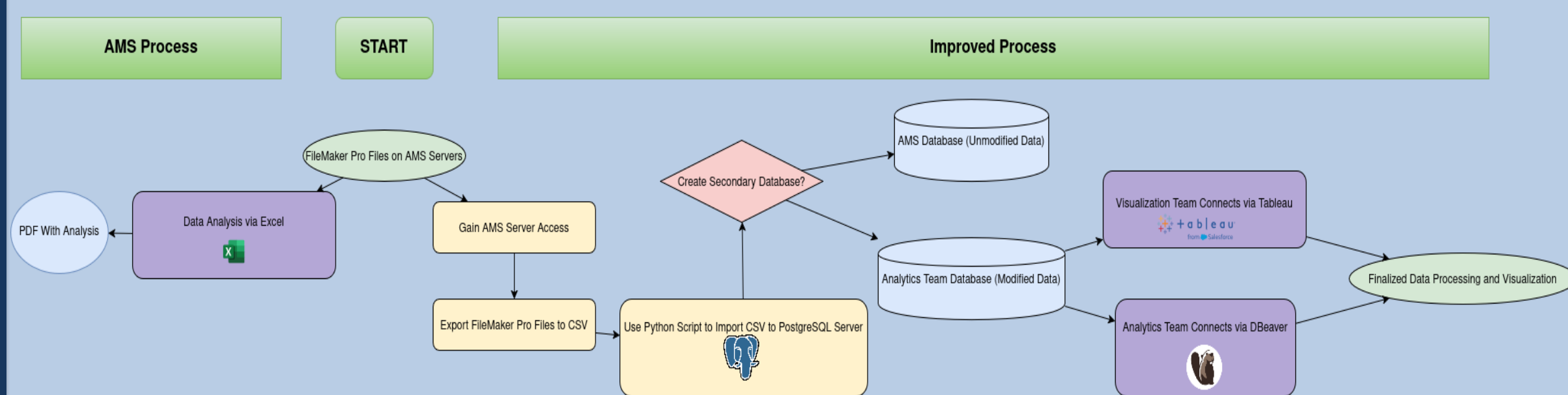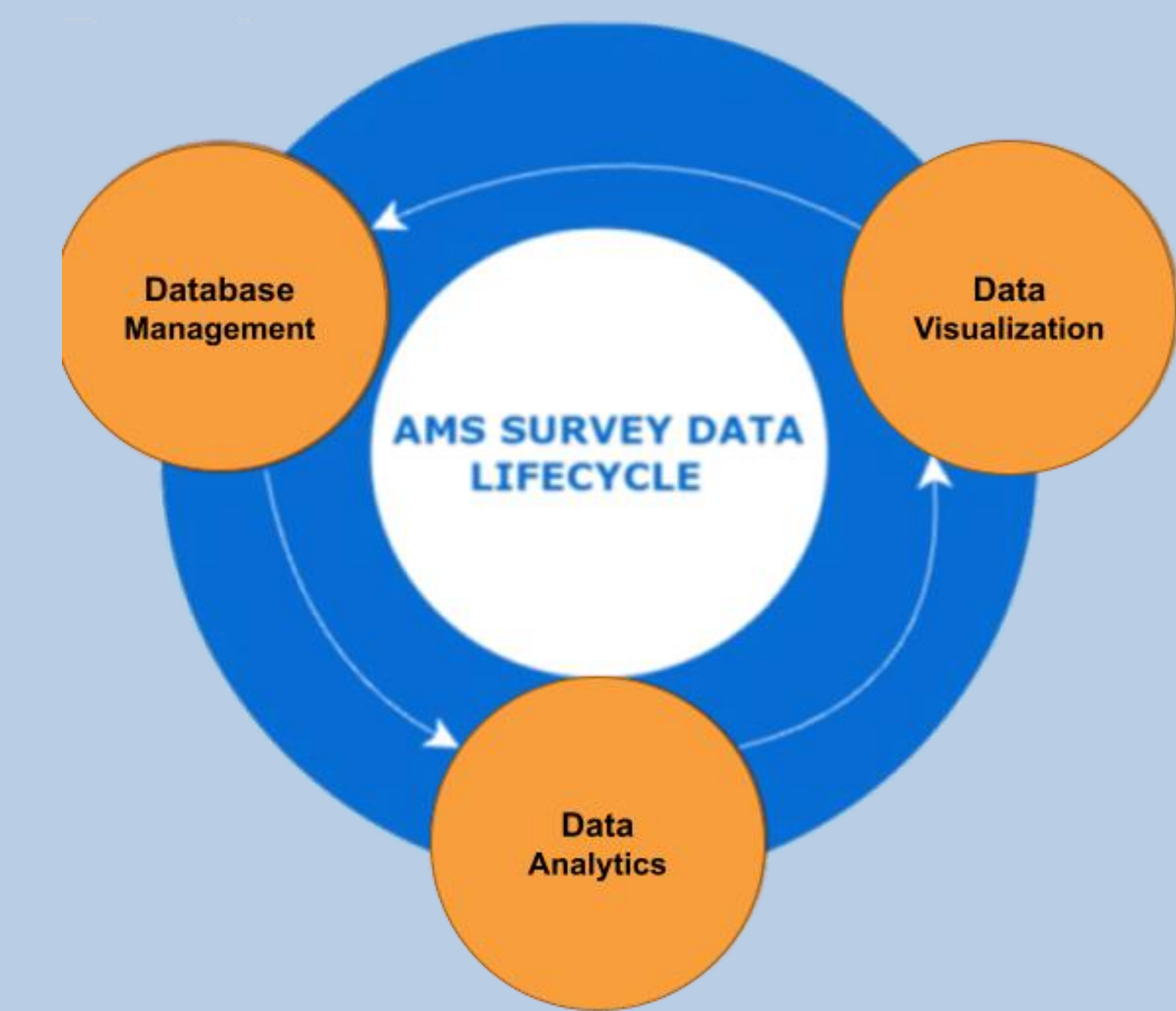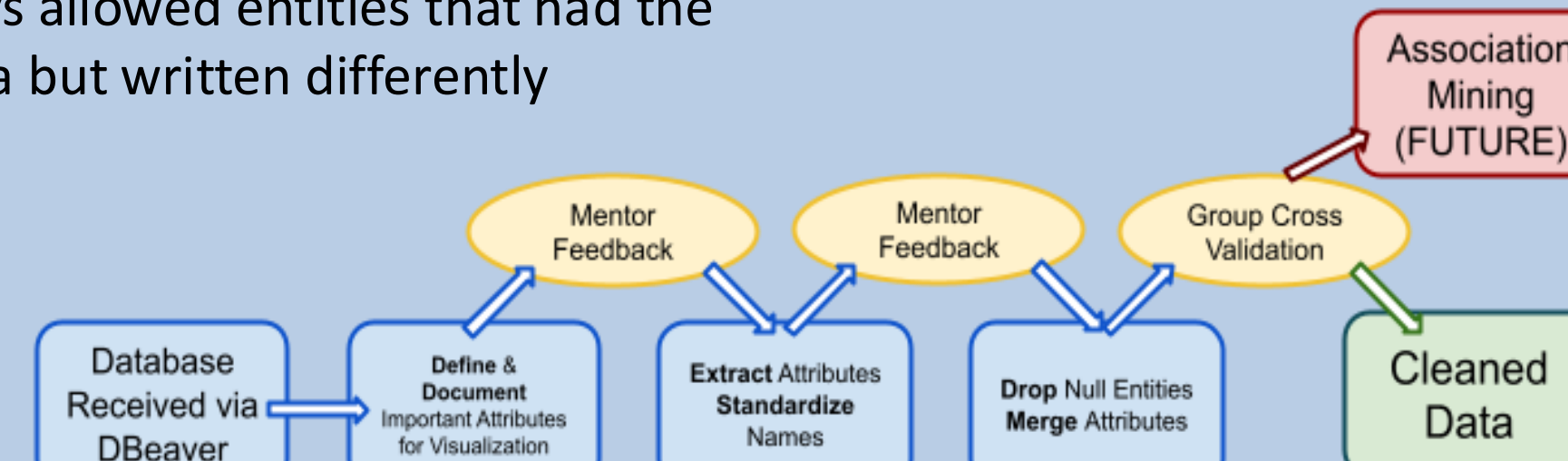

Data Wrangling

WHY?
- For easier visualization process
- Free response surveys allowed entities that had the same categorical data but written differently

HOW?
Flow chart of data wrangling process, resulting in cleaned data and the possibility of association mining for the future.

RESULTS: Before and After data wrangling using anonymous sample data; empty, redundant, or unrelated attributes were deleted as shown in the right.

Gender_Code	MR_Field_EENDR	Salary	Full_Name	Age	Email Reminder Address	Gender_Name
1	1-Algebra	70,000	John Smith	36		Male
2	7-Statistics	10,000	Alice Green	NaN		Female
			Bob Ross			
1	4-Discrete	40,000		49		Male
2	12-Other	NaN		29		Female

Gender_Name	MR_Field_EENDR	Salary	Age
Male	1-Algebra	70,000	36
Female	7-Statistics	10,000	
Male	4-Discrete	40,000	49

AMS SURVEY DATA LIFECYCLE

- Database Management
- Data Visualization
- Data Analytics

Metrics and Preprocessing

WHY?
- Free response surveys allowed entities that had the same categorical data but written differently
- Not enough metrics to describe largely categorical data, especially after cleaning

HOW?
- Built a GitHub repository to store python scripts for preprocessing, analytics/metrics, and data retrieval + conversion
- Designed as not intended to be client facing, but made easier by designing with one run command with configuration flags for future analytics team

RESULTS
- Postgres database data wrangler script that connects to DBeaver and allows metrics code to analyze the data
- Text categorizer that groups similarly worded column data into condensed buckets
- Association mining script will retrieve frequently grouped attributes from table data

Conclusions

In this project, we migrated the AMS data from FileMaker Pro onto a Postgres database in the form of CSVs in which we can export to Tableau to create data visualizations. Throughout this process, we got to experience a professional and structured work environment with real-life non-uniform data and use various tools like Excel, SQL, and Tableau.

Data Visualization

Procedure:

- Exploration of the Data
- Wireframe Design on Figma
- Create Graphs
- Build the Initial Dashboard
- Gathered Feedback
- Finalized the Dashboard
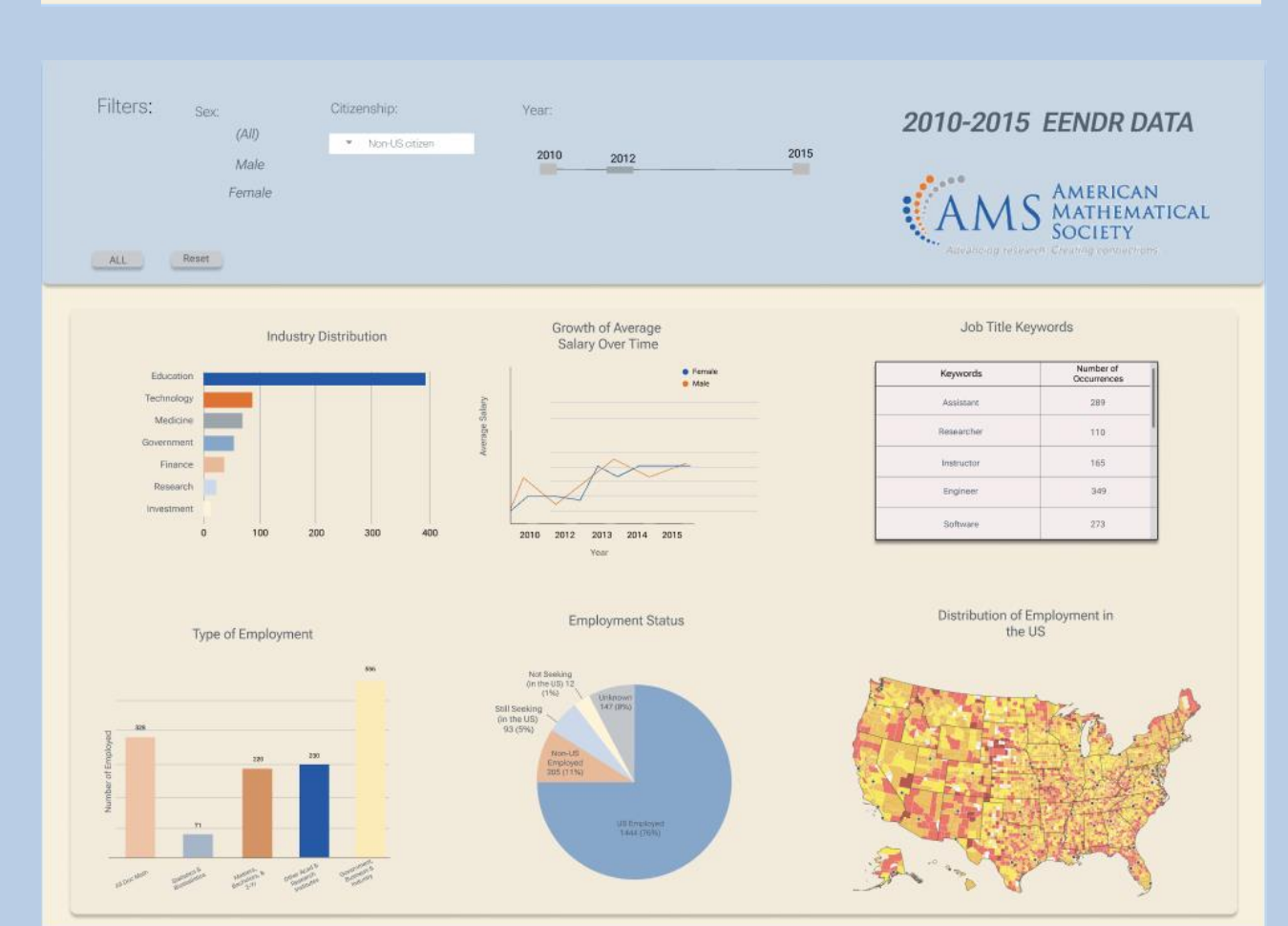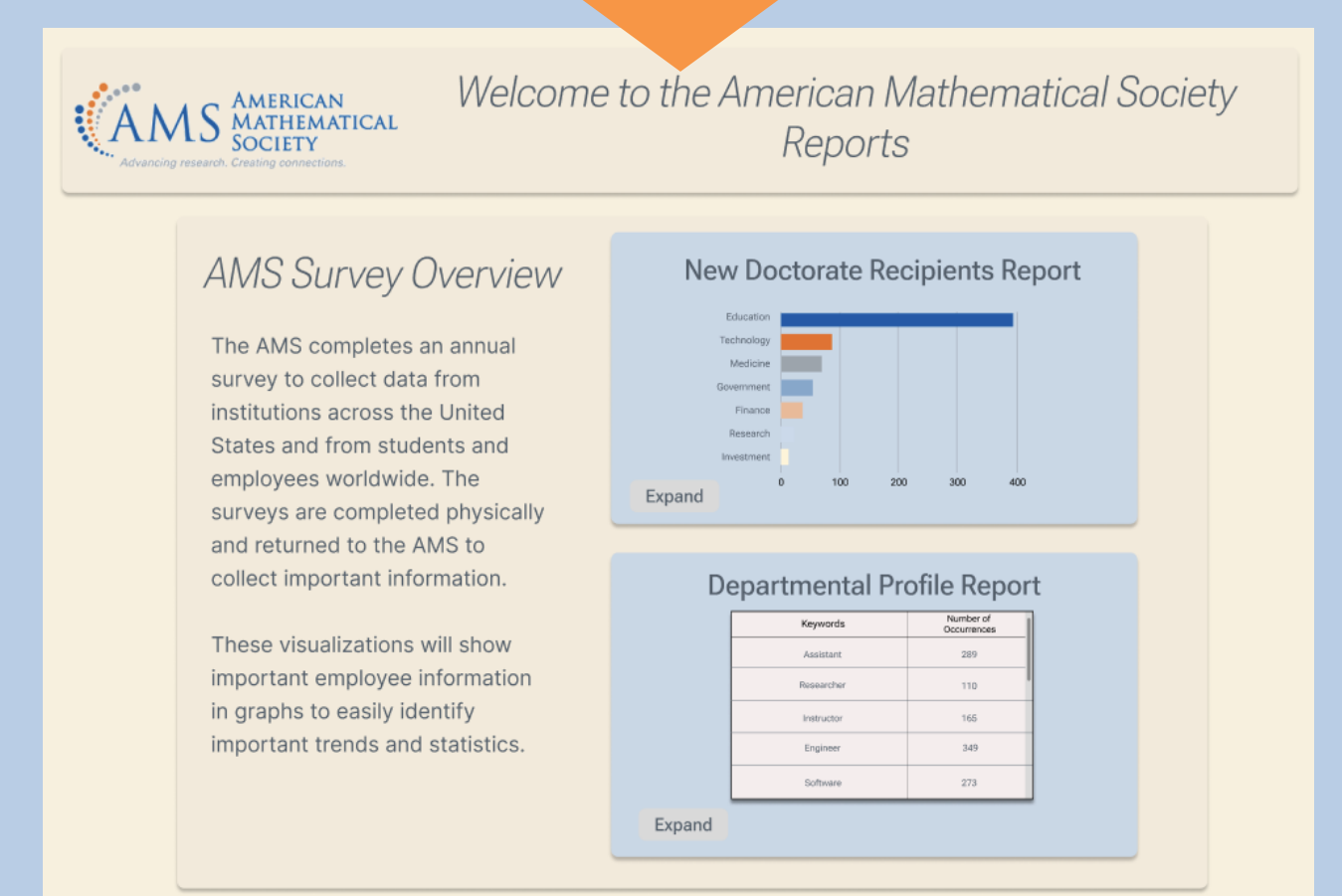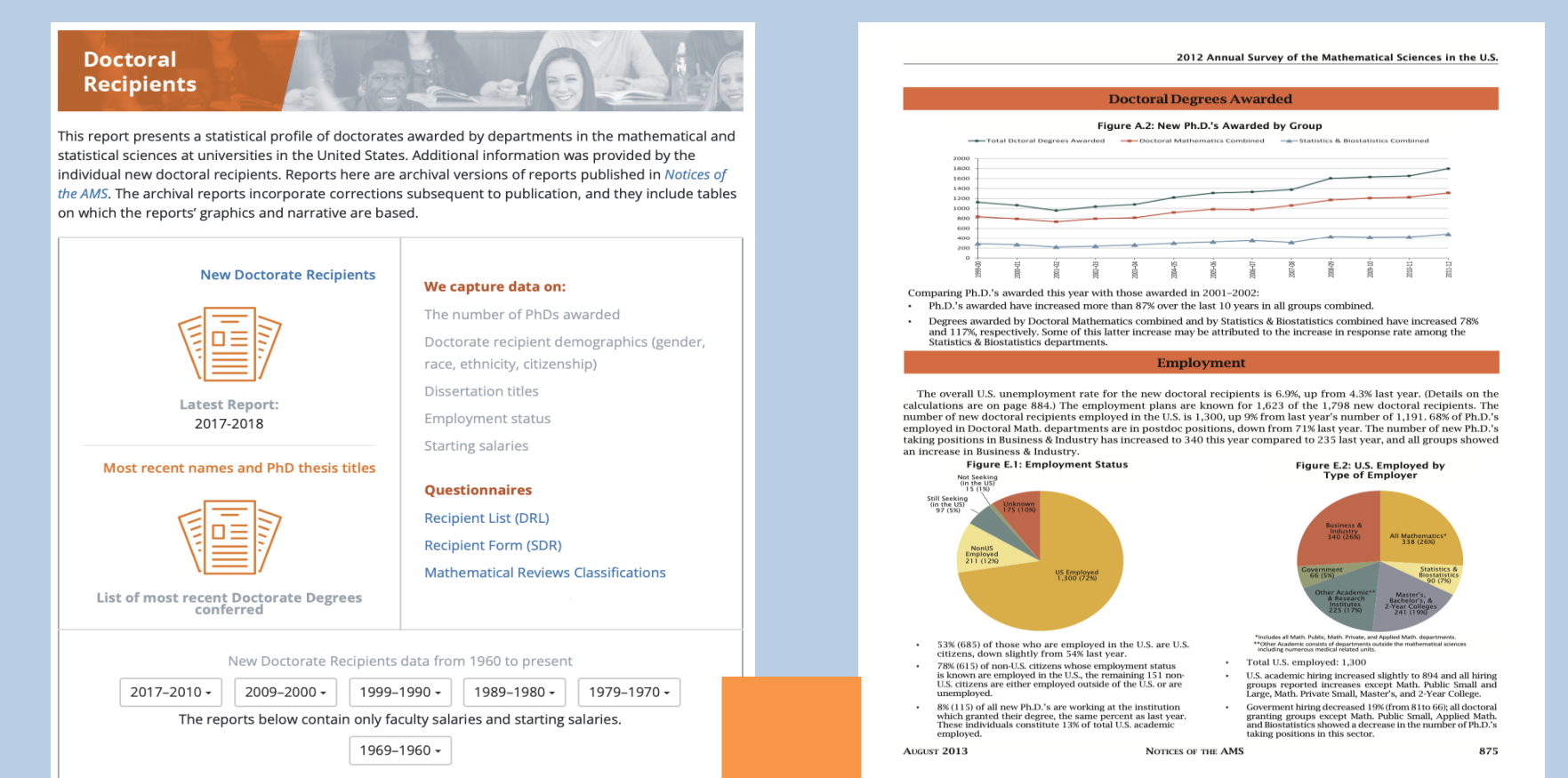
Future Work

- Integrate multi-year data into a centralized system for comprehensive analysis.
- Develop a unified dashboard to display insights across all years.
- Streamline reporting processes for improved accessibility and decision-making.
- Further add additional data mining algorithms that gains distinct insight metrics.
- Reduce server side computations through non-tableau analytics.

Acknowledgements