**Data Source & EDR Process for US Layoff Analysis**

**Sravani yalamarthi**

Saint Peters University

DS-670-HYB2-25SPTR: Capstone: Big Data & Data Science

Professor- ***Reda Mastouri***

**March 09, 2025**

# Abstract

Lay-offs in the workforce have been a recurring phenomenon in different industries due to economic slowdowns, technological advancements, and reorganization in organizations. Knowledge of layoff trends is significant for policymakers, organizations, and researchers in planning counter-strategies for the negative impact of workforce reductions. The current study examines employment separation trends in the Iowa Executive Branch with particular attention focused on the trends in involuntary and voluntary separations. The data, which is extracted through Data.gov, offers extensive records of lay-offs, resignations, and dismissals for different fiscal years. The systematic Extract, Discover, and Refine (EDR) process is used in the current study, with data preprocessing operations such as duplicate removal, categorical encoding, and imputing the missing values. Class imbalance is addressed by the usage of the Synthetic Minority Over-sampling Technique (SMOTE) in order to represent the data in a solid manner. The output of this phase will be used as a basis for exploratory in-depth investigation, detection of trends, and predictive modeling in the later stages of the research.

**Table of Contents**

[Abstract 2](#_Toc192484479)

[1. Introduction 4](#_Toc192484480)

[2. Literature Review 5](#_Toc192484481)

[3. Data Source Identification 7](#_Toc192484482)

[3.2 Data Attributes 8](#_Toc192484483)

[4. Data Preprocessing & EDR Process 9](#_Toc192484484)

[4.1 Data Cleaning 9](#_Toc192484485)

[4.2 Feature Engineering 10](#_Toc192484486)

[4.3 Encoding Categorical Data 10](#_Toc192484487)

[4.4 Addressing Class Imbalance with SMOTE 11](#_Toc192484488)

[4.5 Summary of Preprocessing Outcomes 12](#_Toc192484489)

[Reference 13](#_Toc192484490)

# 1. Introduction

The American workforce has undergone extensive changes with the fluctuating economic conditions, industry demands, and technology advancements. Lay-offs in the workforce, whether voluntary in nature (resignation, retirement) or involuntary (dismissals, lay-offs), impact the workforce, organizations, as well as the economy in general. The employment separation trends are significant in providing relevant information regarding the workforce stability, restructuring at the department level, as well as the impact of the business cycles on job security.

This examines separations in employment in the Iowa Executive Branch, with government data providing the foundation for the examination of historical trends in workforce reductions and layoffs. The data, which is accessible in Data.gov, is a complete record of employee separations beginning in Fiscal Year 2013, including extensive details for job classes, reasons for separations, salary grades, and employment status.

In order to ensure accurate and meaningful results, the research utilizes an Extract, Discover, and Refine (EDR) process with various data preprocessing steps:

* Data cleaning (removing duplicates, filling in missing values, and standardizing the formatting).
* Feature Engineering (extracting other variables for enhancing the interpretability).
* Categorical Encoding (translating textual data to numeric values for analysis).
* Class Balancing: balancing the imbalanced layoff categories with SMOTE.

Through data preparation for in-depth analysis and data cleaning, the aim of this study is to unveil prominent workforce trends and provide significant insights for the benefit of informing workforce planning, policy recommendations, and economic forecasting. The outcome of this exploratory research will be used as the baseline for the subsequent analytical stages in order to facilitate a deeper investigation into the causes of lay-offs as well as workforce mobility.

# 2. Literature Review

1. Bali et al. (2024) examine geospatial employment trends in employment dynamics as well as layoff trends using sentiment analysis of tweets. The study identifies regional employment decrease disparities as well as recruitment trends, providing insight into labor market shifts. The study, using real-time social media data, highlights the application of sentiment analysis in workforce planning. The outcome shows how monitoring the opinions of the general population can enable organizations to forecast labor changes as well as effectively manage employment policies.
2. Eshghi & Astvansh (2024) conduct a meta-analysis of investors in the stock market reacting to layoff announcements with mixed market reactions. The sentiment of investors, the research shows, is influenced by the type of industry, the cause of the layoff, and the state of the economy. While some layoffs trigger short-term increases in the share price due to cost-cutting expectations, others trigger negative market response due to concerns for long-term stability and innovation.
3. Granulo et al. (2025) examine collective lay-offs and offshoring in the light of the social contract, discussing how consumers and stakeholders perceive mass lay-offs. The study finds that consumer confidence is undermined when lay-offs are perceived as unwarranted or profit-driven, which harms brand reputation. Firms engaging in open and ethical restructuring will likely enjoy the support of the general public as well as long-term customer loyalty.
4. Hossen et al. (2023) discuss the factors leading to layoffs in the US high-technology industry with the prime causes being automation, economic downturn, and corporate restructuring. The article highlights how the rapid pace of technological advancements results in displacement, necessitating the reskilling of the workforce. The study states how the effects of layoffs can be lessened by forward-thinking labor policies for the workers as well as the stability of the sector.
5. Joshi (2025) examines the changing role of Agentic Generative AI (GenAI) in workforce education and training in the United States. The research points to the ways in which AI-based automation is transforming the requirements for work, with a greater need for skill flexibility and lifelong learning. The study indicates how AI integration can produce new job opportunities alongside calls for policy interventions to control displacement threats as well as ensure just workforce transitions.
6. Krolikowski & Lunsford (2024) look at the impact of advance layoff notices on total job loss, with the outcome being that advance warnings reduce economic shocks. The study shows that advance layoff announcements give workers the power to seek new jobs, reducing the period of joblessness. The outcome underscores the importance of labor policies ensuring transparency and timely workforce realignments in a bid to prevent economic instability.
7. Liu & Liang (2023) investigate the relationship between permanent job loss and credit card loss forecasting. Their study finds that employment loss significantly increases financial instability, leading to increased levels of default. The research emphasizes the necessity for financial institutions to use employment trends in credit risk models for the purpose of improving forecasting accuracy and minimizing economic risks.
8. Penn & Nezamis (2022) outline the trends in the labor market during the year 2021, with record levels of job openings and quits and record-low layoffs and discharges. The shifts are due to labor shortages caused by the pandemic and shifting work preferences. The outcomes necessitate organizations to realign recruitment and retention strategies to the new workforce realities.
9. Prakash & Sakthivel (2024) elaborate on the use of machine learning algorithms in layoff forecasting and analysis. The study highlights the utilization of predictive analytics for identifying the workers at risk and optimizing workforce planning. Findings show that AI-based decision-making improves the HR strategy, allowing organizations to proactively take measures to minimize layoffs and maximize employee retention.
10. Tomaskovic-Devey et al. (2024) investigate intersectional wage inequality in American workplaces in the public sector during the time of the Great Recession. The research affirms the economic recession disproportionately impacting marginalized groups, which increases wage inequality. The research underscores the need for equal labor policies in addition to targeted interventions to reduce employment and income distribution disparities.

# 3. Data Source Identification

In an attempt to conduct a comprehensive examination of the trends in layoff in the US labor market, the current study utilizes publicly available data gathered by the government. Data used in the current phase is primarily the Iowa Executive Branch Voluntary and Involuntary Employment Separations data obtained through Data.gov. The data provides valuable information about employment trends, reasons for separations, and workforce restructuring over time.

**3.1 Dataset Overview**

The data reports employment separations for the Iowa Executive Branch, including both voluntary separations and involuntary separations, since Fiscal Year 2013, with the data updated bi-weekly. The data reports key employment separation information such as:

• Reasons for Separation – Whether the employee separated voluntarily (for example, retirement, resignation) or involuntarily (for example, lay-offs, dismissals

• Job Classification and Department Data – Includes the names of departments, sub-units, and Equal Employment Opportunity (EEO) categories.

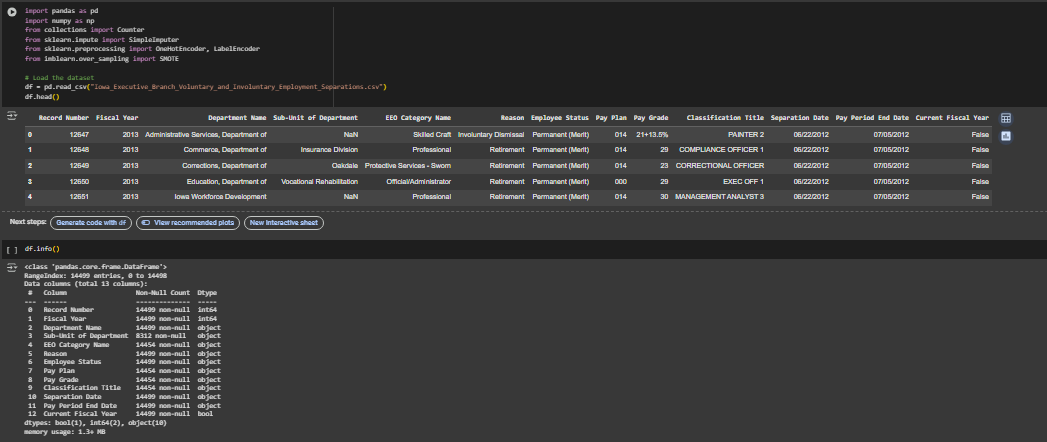
• Salary Grades, Type of Employment, Classification Titles – Includes salary grades, type of employment (permanent or temporary), and classification titles.

• Separation Date – The date of separation and the payment period for the same.

3.2 Data Attributes**:**

|  |  |
| --- | --- |
| **Attribute Name** | **Description** |
| **Record Number** | Unique identifier for each record |
| **Fiscal Year** | Year in which the separation occurred |
| **Department Name** | Department from which the employee separated |
| **Sub-Unit of Department** | Sub-division within the department |
| **EEO Category Name** | Job classification based on Equal Employment Opportunity (EEO) standards |
| **Reason** | Cause of separation (e.g., retirement, dismissal, layoffs) |
| **Employee Status** | Employment type (e.g., permanent, temporary) |
| **Pay Grade** | Employee’s pay grade at the time of separation |
| **Classification Title** | Employee’s job title |
| **Separation Date** | Date when the employee left the job |
| **Pay Period End Date** | Pay period associated with the separation |
| **Current Fiscal Year** | Boolean flag indicating if the record belongs to the current fiscal year |

# 4. Data Preprocessing & EDR Process

Data preprocessing is necessary to ensure data quality, integrity, and usability for future analysis. The Extract, Discover, and Refine (EDR) process applied in this study consists of various steps including data cleaning, feature engineering, categorical encoding, and handling class imbalances.

## 4.1 Data Cleaning

The data underwent several cleaning procedures to enhance precision and consistency:

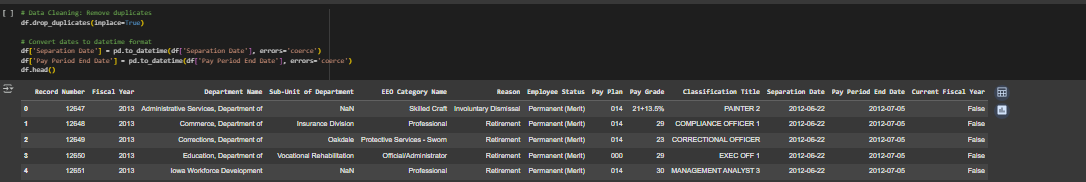
• Removal of duplicates: Eliminated duplicate records to prevent data distortion.

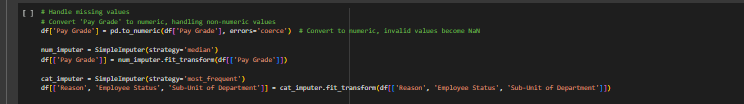
• Date Formatting: Converted Separation Date and Pay Period End Date to datetime type for easier chronological comparison.

• Missing Value Handling:

Numeric Columns: Numerical columns with missing values (for instance, Pay Grade) were imputed with the median in order to minimize skewness.

Missing values in the categorical variables (Employee Status, Reason, Sub-Unit of Department) were replaced with the most frequent category in each group.

****

****

## 4.2 Feature Engineering

To derive additional insights, new features were created from existing attributes:

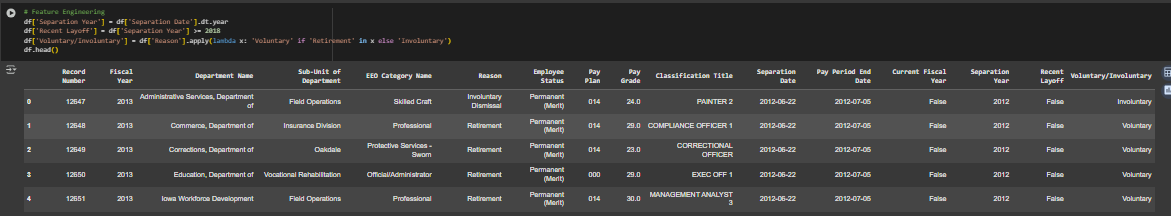
• Separation Year: Derived from the Separation Date in order to examine year-by-year trends in separations.

• Recent Layoff Flag: Included a Boolean flag for layoffs subsequent to the year 2018, reflecting the latest labor market trends.

• Separation Type Classification

Voluntary Separations: Transfers, retirements, and

* Involuntary Separations: Layoffs, dismissals, and terminations.
* They make the data more interpretable and easier for predictive modeling.

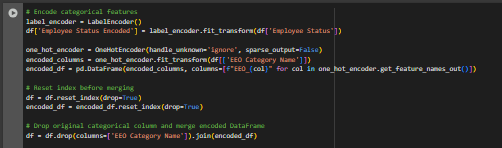
****

## 4.3 Encoding Categorical Data

Categorical variables were converted to numerical values for statistical analysis as well as training the machine learning model:

• Label Encoding: Used for Employee Status to provide numerical values for the type of employment.

• One-Hot Encoding: Used for categorical variables such as EEO Category Name and Reason, converting them into binary columns for the purpose of increased model compatibility.

****

## 4.4 Addressing Class Imbalance with SMOTE

The data revealed a wide class imbalance, most significantly in the Reason column, with voluntary separations heavily outnumbering the involuntary separations. To offset this imbalance:

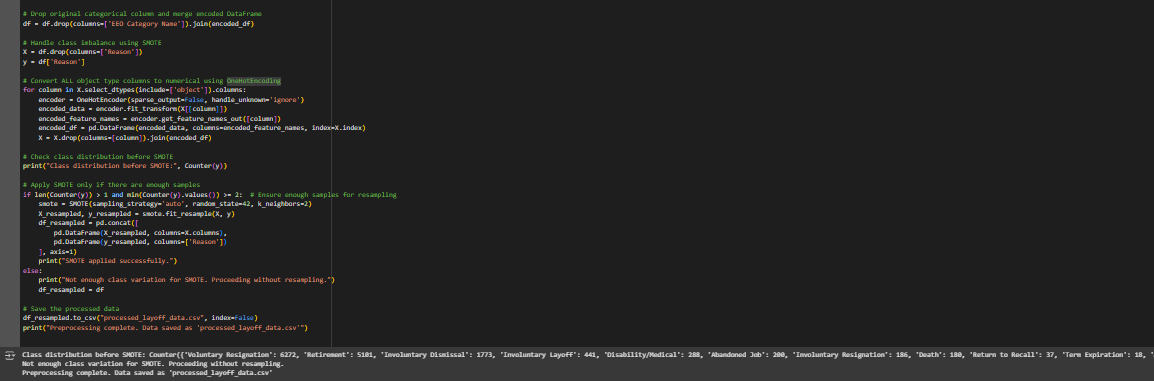
1. Applied One-Hot Encoding to transform categorical variables into numeric variables.

2. Utilized Applied Synthetic Minority Over-sampling Technique (SMOTE) for balancing classes.

3. Utilized k-Nearest Neighbors with k=2 for enhanced synthetic

4. Verified Class Distribution before and after resampling for data quality.

SMOTE guarantees the equal treatment of dismissals and layoffs, leading to a balanced data set for predictive modeling.

****

## 4.5 Summary of Preprocessing Outcomes

The data were enhanced after the completion of the EDR process as

Data consistency – Missing values have been imputed, duplicate records eliminated, and date formatting normalized.

Feature Engineering – New variables, whether categorical or numeric, created to provide greater analytical insight.

Class Balancing – Utilizing SMOTE to prevent model bias due to underrepresented layoff categories.

Processed Data Set for Analysis – The processed and cleaned data was saved for future statistical and machine learning purposes as processed\_layoff\_data.csv.

# Reference

Bali, R., Sharma, A., Mala, S., & Malhan, Y. (2024). Modeling the Geospatial Trend Changes in Jobs and Layoffs by Performing Sentiment Analysis on Twitter Data. *International Journal of Performability Engineering*, *20*(2).

Eshghi, K., & Astvansh, V. (2024). Stock investors' reaction to layoff announcements: A meta‐analysis. *Human Resource Management Journal*, *34*(3), 792-809. <https://onlinelibrary.wiley.com/doi/full/10.1111/1748-8583.12532>

Granulo, A., Kranzbühler, A. M., Fuchs, C., & Puntoni, S. (2025). Collective Layoffs and Offshoring: A Social Contract Account. *Journal of Consumer Research*, ucaf001. <https://doi.org/10.1093/jcr/ucaf001>

Hossen, M. N., Mollah, M. A., Lipy, N. S., Hossain, G. M. S., & Rahman, M. S. (2023). Factors affecting layoff in high-tech industry: Evidence from USA. *International Journal of Information, Business and Management*, *15*(1), 1-13. <https://www.proquest.com/openview/dff61e6a756dd2a2a006b8407833566d/1?pq-origsite=gscholar&cbl=2032142>

<https://openurl.ebsco.com/EPDB%3Agcd%3A8%3A18249768/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Agcd%3A175379479&crl=c&link_origin=scholar.google.com>

<https://www.jstor.org/stable/48716861>

Joshi, S. (2025). The Transformative Role of Agentic GenAI in Shaping Workforce Development and Education in the US. *Iconic Research And Engineering Journals*, *8*(8), 199-206.

Krolikowski, P. M., & Lunsford, K. G. (2024). Advance layoff notices and aggregate job loss. *Journal of Applied Econometrics*, *39*(3), 462-480. <https://doi.org/10.1002/jae.3032>

Liu, Z., & Liang, H. (2023). Permanent layoff and consumer credit card loss forecasting. Managerial Finance, Vol. 49 No. 5, pp. 789-807. <https://doi.org/10.1108/MF-02-2022-0085>

Penn, R., & Nezamis, E. (2022). Job openings and quits reach record highs in 2021, layoffs and discharges fall to record lows. *Monthly Labor Review*.

Prakash, P., & Sakthivel, V. (2024, February). Layoffs Analysis and Prediction Using Machine Learning Algorithms. In International Conference on Communications and Cyber Physical Engineering 2018 vol 1096 (pp. 535-543). Singapore: Springer Nature Singapore <https://doi.org/10.1007/978-981-99-7137-4_53>

Tomaskovic-Devey, D., Kerrissey, J., Rainey, A., & Boutcher, S. A. (2024). Intersectional Earnings Inequalities in US Public Sector Workplaces and the Great Recession. *Socius*, *10*, 23780231241257778. <https://doi.org/10.1177/23780231241257778>