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

**Likhita Alla**

Saint Peters University

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

Professor- ***Reda Mastouri***

**March 09, 2025**

# Abstract

The US job market has experienced significant changes due to various economic, technological, and policy factors, leading to shifts in employment trends. The current study is focused on the investigation of layoffs in the Iowa Executive Branch using employment separation data obtained from government sources. The primary purpose is to ascertain trends in involuntary and voluntary separations, examine the underlying causes for the layoffs, and apply the use of machine learning to increase data understanding. The study employs data preprocessing steps such as duplicate removal, imputing missing values, and categorical encoding, in addition to the application of the Synthetic Minority Over-sampling Technique (SMOTE) for balancing classes. The preprocessed data enable systematic understanding of workforce reductions, which inform strategic decision-making by policymakers, HR professionals, and labor economists. The current report presents a comprehensive overview of data sources, preprocessing steps, and analytical methodologies applied in Phase 1 of the study.

**Table of Contents**

[Abstract 2](#_Toc192484413)

[1. Introduction 4](#_Toc192484414)

[2. Literature Review 5](#_Toc192484415)

[3. Data Source Identification 7](#_Toc192484416)

[3.1 Dataset Overview 7](#_Toc192484417)

[3.2 Data Attributes 7](#_Toc192484418)

[4. Data Preprocessing & EDR Process 8](#_Toc192484419)

[4.1 Data Cleaning 8](#_Toc192484420)

[4.2 Feature Engineering 9](#_Toc192484421)

[4.3 Encoding Categorical Data 10](#_Toc192484422)

[4.4 Addressing Class Imbalance with SMOTE 10](#_Toc192484423)

[4.5 Results & Processed Data 11](#_Toc192484424)

[Reference 12](#_Toc192484425)

# 1. Introduction

Workforce trends in the US have been drastically altered in the recent past by economic recessions, automation, restructuring by organizations, and shifts in labor legislation. Voluntary and involuntary lay-offs are at the center in the creation of employment trends and have significant implications for workers as well as organizations. Policymakers, organizations, and researchers should be aware of these trends in order to counter the economic impact as well as to increase workforce planning.

This report presents an overview of separations in the Iowa Executive Branch using publicly available data for voluntary and involuntary separations. The data, which is derived from Data.gov, comprises records for Fiscal Year 2013 onwards, with detailed information for the causes of separations, job categories, employment trends by departments, and payment structures.

In order to make meaningful inferences, the current research employs an Extract, Discover, and Refine (EDR) approach with strict preprocessing steps such as duplicate removal, handling of missing values, and encoding of categorical variables to maintain data integrity. Moreover, to counter the inherent imbalanced nature of the causes of layoff, SMOTE-based resampling is applied, which improves the robustness of the resulting analytical models.

The study aims to gain a deeper understanding of layoff trends, with the goal of providing valuable information for stakeholders including government agencies, labor economists, and HR professionals. The findings can be applied to inform policy interventions, workforce planning initiatives, and predictive modeling for upcoming changes in employment. Phase 1 of the study lays the groundwork by structuring and preprocessing the data for the higher-level analytics in the subsequent phases.

# 2. Literature Review

**Hamouche, Kammogne, and Merkouche (2023)** examine the impact of crisis-induced career shocks among the workforce in terms of job insecurity, redundancy, and felt employability. The study identifies inequalities in the workforce by education, gender, and ethnicity with the marginalized groups having higher job vulnerability. The findings show long-term employment insecurity and altered career ambitions due to career shocks.

**Liu and Liang (2023)** investigate the relationship between permanent layoff and credit card loss forecasting among consumers. Their research shows how financial instability increases in the aftermath of layoffs, influencing consumer spending and debt payment behavior. The study, through its predictive models, shows how the economic ripple effects of layoffs require financial planning and risk mitigation measures.

**Bhat and Sarkar (2024)** examine the domino effect of the recession on economic and psychological capital. They outline how loss of employment results in decreased productivity, economic recession, and mental distress for the affected workers. The study provides insight into long-term recovery challenges, with the contribution of policy interventions and workforce resilience measures in mitigating recessionary effects.

**Ridho and Azizah (2022)** investigate mass lay-offs in start-ups by adopting a mixed-method approach with structural equation modeling. The most prominent factors are found by the study to be market volatility, poor financial planning, and investors' confidence as the main reasons for lay-offs. The study highlights the unique problems of start-ups wherein financial instability leads to uncertain lay-offs, affecting job security.

**Hossen et al. (2023)** investigate lay-offs in the US high-technology industry, with driving factors for job reductions analyzed. The study highlights the role of automation, international outsourcing, and economic recession in influencing employment trends. Findings show corporate efficiency is boosted by lay-offs, although their effects are felt in morale among the workforce as well as long-term industry stability. The study advocates for anticipatory workforce adjustment measures.

**Prakash and Sakthivel (2024)** focus specifically on the use of machine learning algorithms in the forecasting of layoffs. The paper demonstrates how AI-based models help in improving workforce reduction forecasting, enabling organizations to design anticipatory strategies. The article underscores the strength of data-driven decision-making in planning for people, with particular emphasis placed on how predictive analytics can minimize the ill effects of layoffs.

**Lazzari, Alvarez, & Ruggieri (2022)** examine employee turnover intentions using predictive analytics. Their research identifies the role of job satisfaction, organizational commitment, and work-life balance in turnover decisions. The study, through the use of machine learning algorithms, presents employee attrition warning signals to organizations, enabling them to implement preventive actions for employee retention. The study shows data-based HR policies can be effective in reducing turnover.

**Abdolmaleki et al. (2024)** discuss employee turnover in the construction industry, the primary causes of which are job insecurity, working conditions, and constraints in career growth. The article states the long-term consequences of turnover, including cost overruns and project delays. The study emphasizes the importance of adopting good workforce management practices and identifies future directions for enhancing employee retention in the construction sector.

**Yahia, Hlel, & Colomo-Palacios (2021)** examine the use of big data and deep data in predicting employee attrition. The article shows how AI-based analytics can provide real-time workforce trends for the purpose of employee retention. The article highlights the necessity of advanced data processing in the identification of high-risk workers and advises the use of predictive modeling for the improvement of the efficiency of HR decision-making.

**Saba (2024)** investigates the impact of lay-offs on the performance of U.S. technology companies. The study indicates that lay-offs raise short-term profitability, but at the expense of long-term issues such as reduced innovation, employee discontent, and damage to the company reputation. The study calls for strategic workforce planning to offset the negative effects, with the requirement for balanced approaches to the process of reducing the workforce.

# 3. Data Source Identification

The data for the current study are derived from Data.gov, specifically the Iowa Executive Branch Voluntary and Involuntary Employment Separations data. The data track historical employee separations in the Iowa Executive Branch, including workforce trends as well as trends in lay-offs. The data are updated bi-weekly and include extensive information about various aspects of employment separation, including the names of departments, job classes, pay grades, and the causes of separation.

## 3.1 Dataset Overview

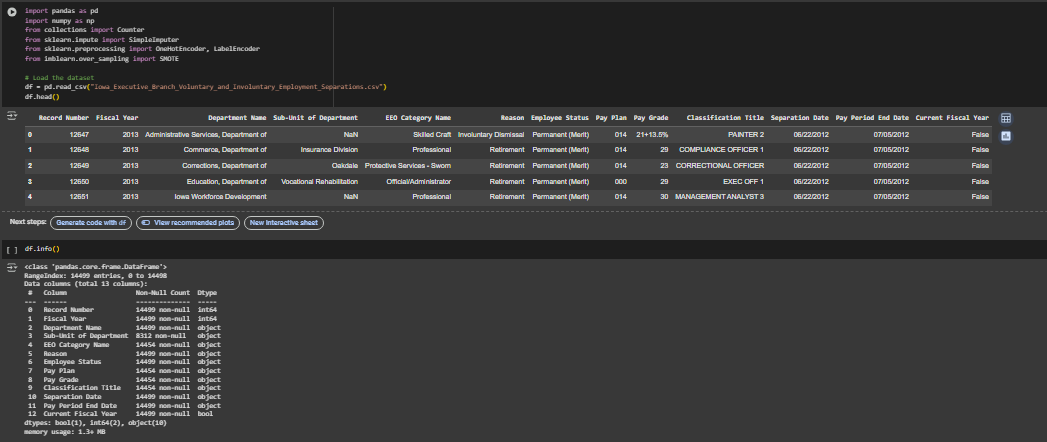
The data consists of systematic employment records of involuntary and voluntary separations. It is an excellent tool for the study of workforce reductions, turnover by jobs, and layoff trends by departments.

3.2 Data Attributes**:**

* **Record Number**: Unique identifier for each record
* **Fiscal Year**: Year in which separation occurred
* **Department Name**: Department from which the employee separated
* **Sub-Unit of Department**: Sub-division of the department
* **EEO Category Name**: Job classification based on Equal Employment Opportunity (EEO) standards
* **Reason**: The 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**: Job title
* **Separation Date**: Date of employee separation
* **Pay Period End Date**: Pay period end date associated with separation
* **Current Fiscal Year**: Boolean flag indicating whether the record is from the current fiscal year

# 4. Data Preprocessing & EDR Process

Several preprocessing steps were carried out in order to ensure data quality and enhance the accuracy of the analytics. This section outlines the cleaning, transformation, and data refining in line with the Extract, Discover, and Refine (EDR) process.



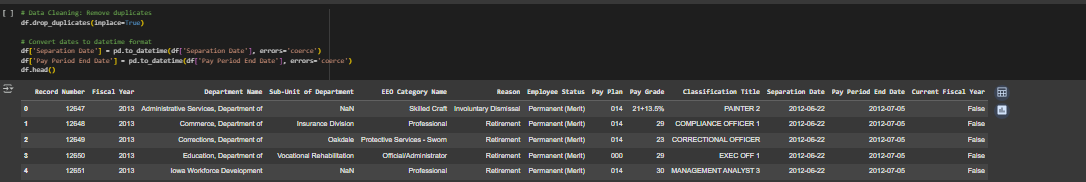
## 4.1 Data Cleaning

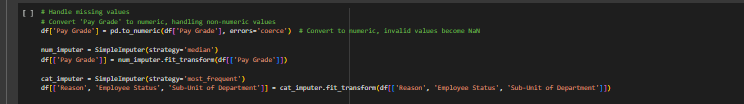
In preparation for the data to be analyzed, the following cleaning procedures were used:  
***Duplicate Removal*** – Eliminated duplicate records to prevent redundancy.

***Date Conversion*** – Converted the fields of "Separation Date" and "Pay Period End Date" to datetime for time-based analysis.

***Handling Missing Values:***

Numerical Columns (for example, Pay Grade) – The missing values were replaced with the median in order to maintain consistency.  
Categorical Columns (i.e., Reason, Employee Status, Sub-Unit of Department) – The mode (most frequent value) in each column was used to impute the missing values.

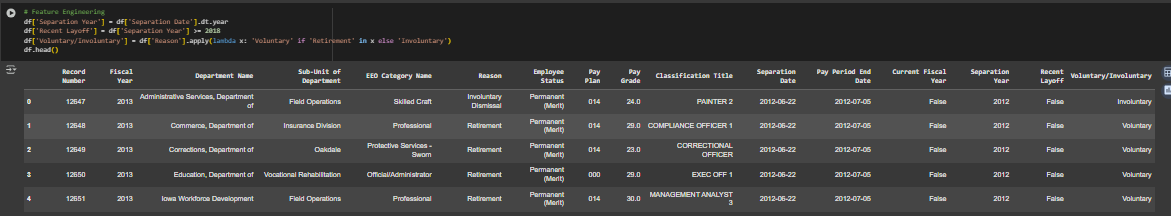
****

****

## 4.2 Feature Engineering

Additional features were created in order to enrich the data for analysis:  
***Separation Year*** – Selected the year column "Separation Date" to look at year-wise trends.  
***Recent Layoff Flag*** – A Boolean flag for separations after the year 2018 to evaluate the most up-to-date employment trends.

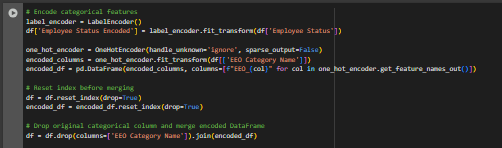
***Categorized Separations***: The separations were categorized as Voluntary (for example, resignation, retirement) and Involuntary (for example, dismissal, layoff

****

## 4.3 Encoding Categorical Data

Since the statistical models and the machine learning algorithms require numerical data, the categorical variables were converted

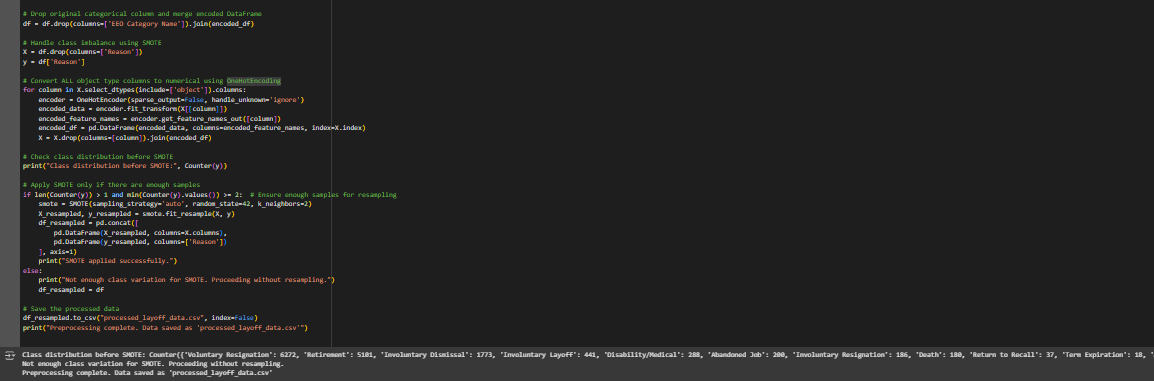
* **Label Encoding** – Applied in the column "Employee Status" for converting categorical variables into numeric values.
* **One-Hot Encoding** – Used for "EEO Category Name" and other categorical features for enhancing the compatibility of the model.

****

## 4.4 Addressing Class Imbalance with SMOTE

The data showed a stark disproportion between the involuntary and the voluntary separations, necessitating the use of Synthetic Minority Over-sampling Technique (SMOTE) for balancing the classes:

1. Converted categorical data into numeric data by using One-Hot Encoding.
2. Utilized SMOTE to create additional records in the minority split category (for example, layoffs).
3. Used k-nearest neighbors (k=2) to generate synthetic samples with maintained distribution of the dataset.
4. Verified class distribution prior to and after resampling to ensure improvements in balance.

****

## 4.5 Results & Processed Data

After preprocessing, the cleaned and sorted data was saved as "**processed\_layoff\_data.csv**." The data is now suitable for exploratory data analysis (EDA), visualization, and predictive modeling in the next phases.

# Reference

Abdolmaleki, G., Naismith, N., Ghodrati, N., Poshdar, M., & Babaeian Jelodar, M. (2024). An analysis of the literature on construction employee turnover: drivers, consequences, and future direction. Construction Management and Economics, 42(9), 822-846. <https://doi.org/10.1080/01446193.2024.2337084>

Bhat, R., & Sarkar, P. (2024). Recession: A Conceptual Case Study Of Its Domino Effect On Economic And Psychological Capital. *Indonesian Journal of Management Studies*, *3*(1), 44-58. <https://doi.org/10.53769/ijms.v3i1.856>

Hamouche, S., Kammogne, C. L., & Merkouche, W. (2023). Workforce's crisis-induced career shock, career preferences, job insecurity, layoff and perceived employability: examining variations based on gender, education level and ethnic origin. *Higher Education, Skills and Work-Based Learning*, *13*(2), 264-282. <https://doi.org/10.1108/HESWBL-05-2022-0101>

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, Vol. 15, Iss. 1, (Feb 2023): 1-13. <https://www.proquest.com/openview/dff61e6a756dd2a2a006b8407833566d/1?pq-origsite=gscholar&cbl=2032142>

Lazzari, M., Alvarez, J. M., & Ruggieri, S. (2022). Predicting and explaining employee turnover intention. International Journal of Data Science and Analytics, 14(3), 279-292.  <https://doi.org/10.1007/s41060-022-00329-w>

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>

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>

Ridho, W. F., & Azizah, N. (2022). Factor Analysis of the Phenomenon of Mass Layoffs At Startups: Mixed Method Approach With Structural Equation Modeling. *Jurnal Mebis*, *7*(2), 195-208. <https://doi.org/10.33005/mebis.v7i1.373>

Saba, Z. (2024). Layoffs and corporate performance: evidence based on the US tech industry. Journal of Economics and Finance, 48(3), 644-667. <https://doi.org/10.1007/s12197-024-09673-y>

Yahia, N. B., Hlel, J., & Colomo-Palacios, R. (2021). From big data to deep data to support people analytics for employee attrition prediction. Ieee Access, 9, 60447-60458. [10.1109/ACCESS.2021.3074559](https://doi.org/10.1109/ACCESS.2021.3074559)