INDEX

PROJECT ONE

ENTERTAINER DATA ANALYSIS

INTRODUCTION

The Entertainer Data Analysis project aims to provide a comprehensive examination of the entertainment industry by analyzing data related to various entertainers, including actors, musicians, and other public figures. The project leverages data analytics to gain insights into trends, popularity, and the impact of entertainers on culture and society.

The objective of this project is to collect data on entertainers and using Tableau to create a detailed dashboard, this project will analyze their performance and impact. The dashboard will visualize trends, popularity metrics, and audience demographics. This will provide stakeholders with comprehensive insights into the entertainment industry.

PROBLEM STATEMENT

Normal life can be stressful, and people need to relax. Being entertained by others is a wonderful way to take some time out of life. It can reduce stress and make life's issues easier to face. The media and entertainment industry consists of film, television, radio and print. These segments include movies, TV shows, radio shows, news, music, newspapers, magazines, and books. The entertainment industry is a group of sub-industries devoted to entertainment. The entertainment industry is used to describe the mass media companies that control the distribution and manufacture of mass media entertainment.

DATASET INFORMATION

The data was split into three parts:

- Entertainer Basic Info,
- Entertainer Breakthrough Info, and
- Entertainer Last Work Info.

→ Combined the datasets into a single file named “Entertainer Final Data”.

→ Also, Given data was not sufficient for analysis part , more information was required so added other information as well.

→ Number of awards and nominees won by them, awards from breakthrough performances were added to the data. The data is in the form of numeric and alphabetic values.

→ Additional data was obtained from IMDb’s official website.

→ Source: IMDB

EXPECTED OUTCOMES

- Comprehensive reports on trends and popularity in the entertainment industry.
- Insights into the cultural and social impact of entertainers.
- Predictive models for future trends and rising stars.
- Visualizations and dashboards for stakeholders to easily understand and interact with the data.

ARCHITECTURE:

Data Source **Connecting data source to Tableau**

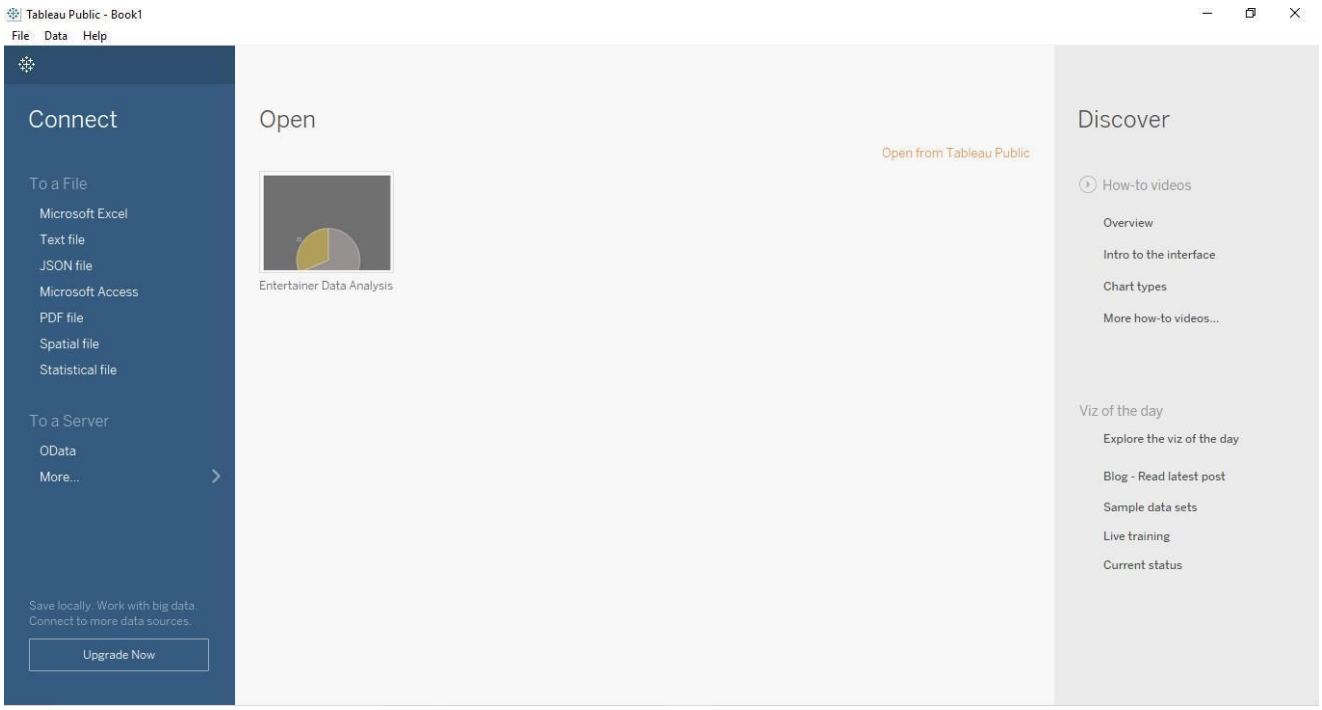
- The architecture of the Entertainer Data Analysis project involves collecting and integrating data from sources like IMDBpy, local MySQL databases, and CSV files.
- After Excel, I pushed this dataset in Pandas (Python library) pandas will manipulate the tables and merged them into one table. Then converted the table into excel file to push into Tableau server to make a dashboard. A connection is then established between Tableau and the database to import the "Entertainer Final Data" file.
- In Tableau, interactive dashboards are created, utilizing various visualization tools and customization options.
- Screen of **Tableau** desktop, client and various charts and **dashboard** screen of Tableau are present at client side.
- After selecting the data in the form of rows and columns it will go inside the tableau server. In the tableau server, it understands the query and generates the best recommended charts based on selected data and returns it into the tableau screen.

- If a client is not **satisfied** with the result, he/she has to select data accordingly otherwise make required changes to show the expected result.

CONNECT DATA WITH TABLEAU AND DEPLOYMENT:

- First of all, open Tableau Public on your desktop. At the first screen, it will ask you to connect your files from various sources like MS Excel, SQL Server, Tableau Server etc.

- First screen of Tableau looks like:

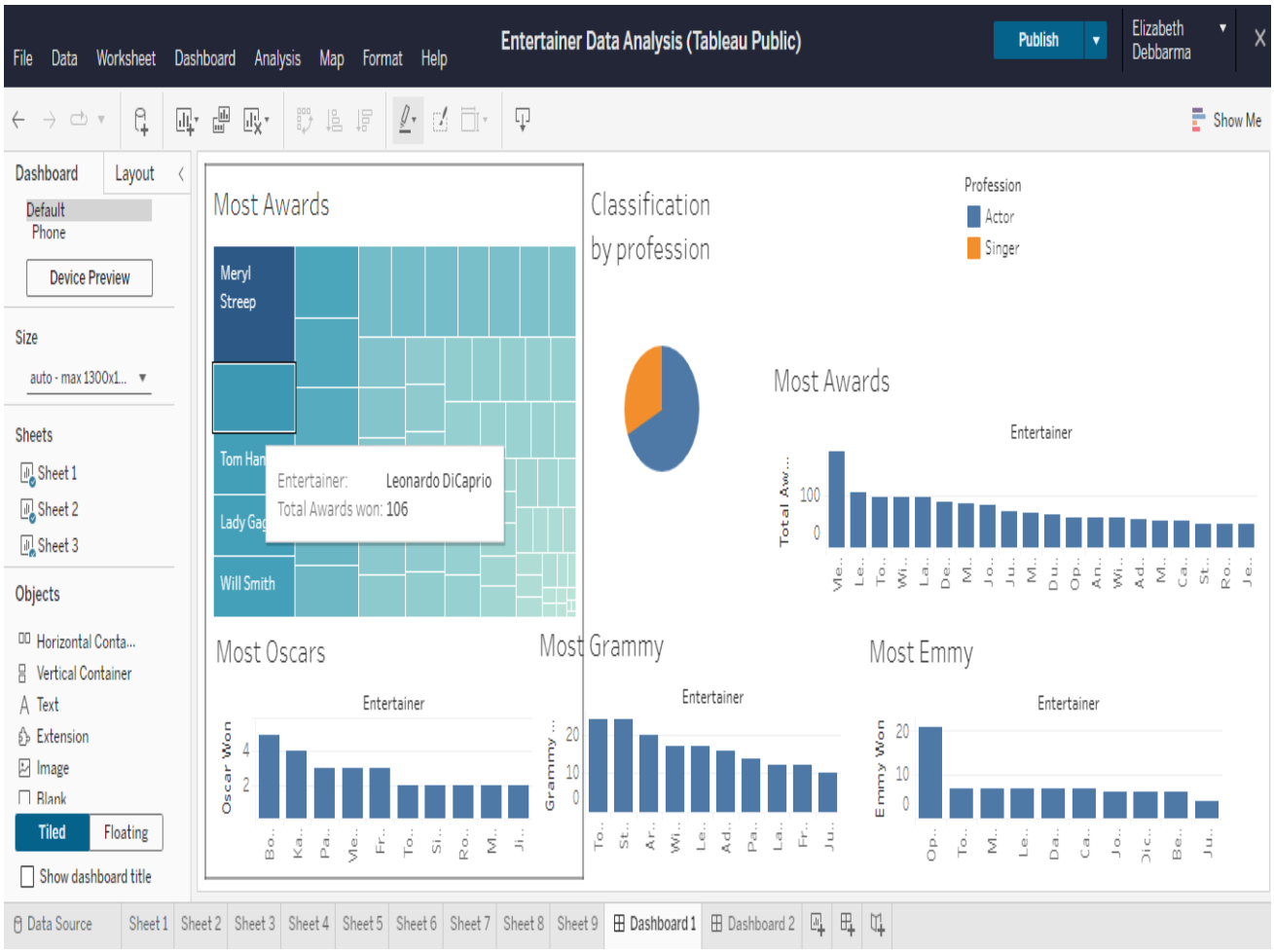


- Make sure the internet connection is connected well while working with tableau, otherwise it will show the error.

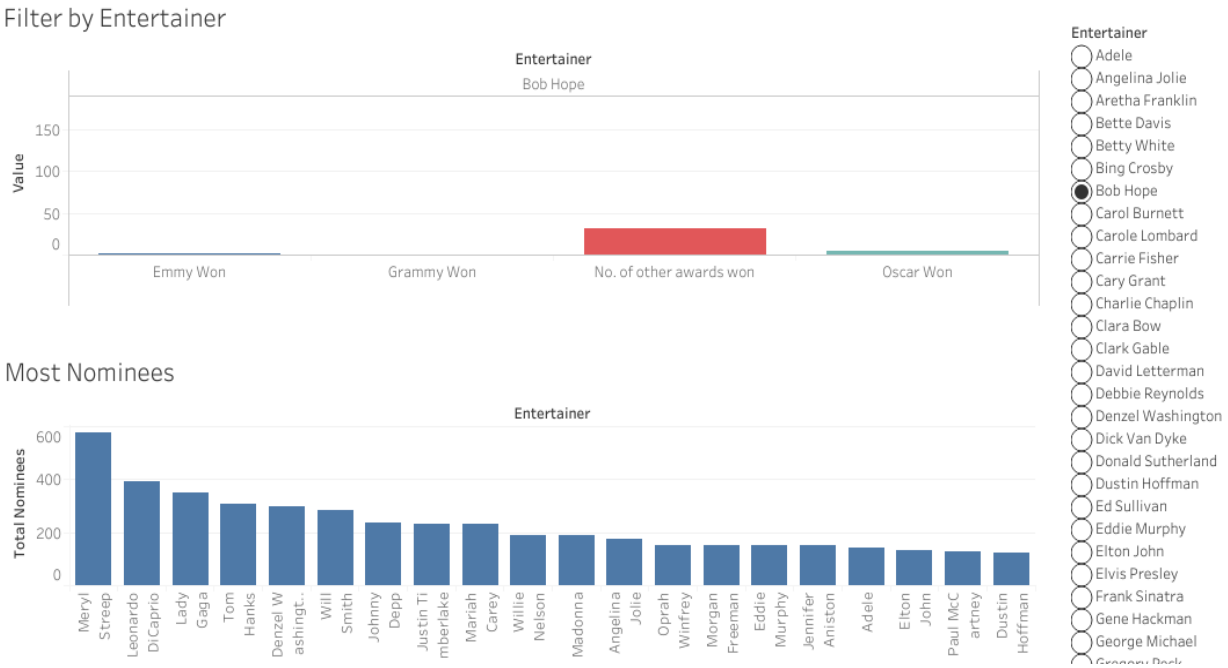
- After completion of work, you can simply press ctrl + s or save it from the file menu. It will let you to tableau public's website and ask you for signing in. After sign-in, your work will be saved on tableau's website. There, all can see the work.

- In **Tableau** we can make professional dashboards as you can in the dashboards provided to you on Tableau public server.

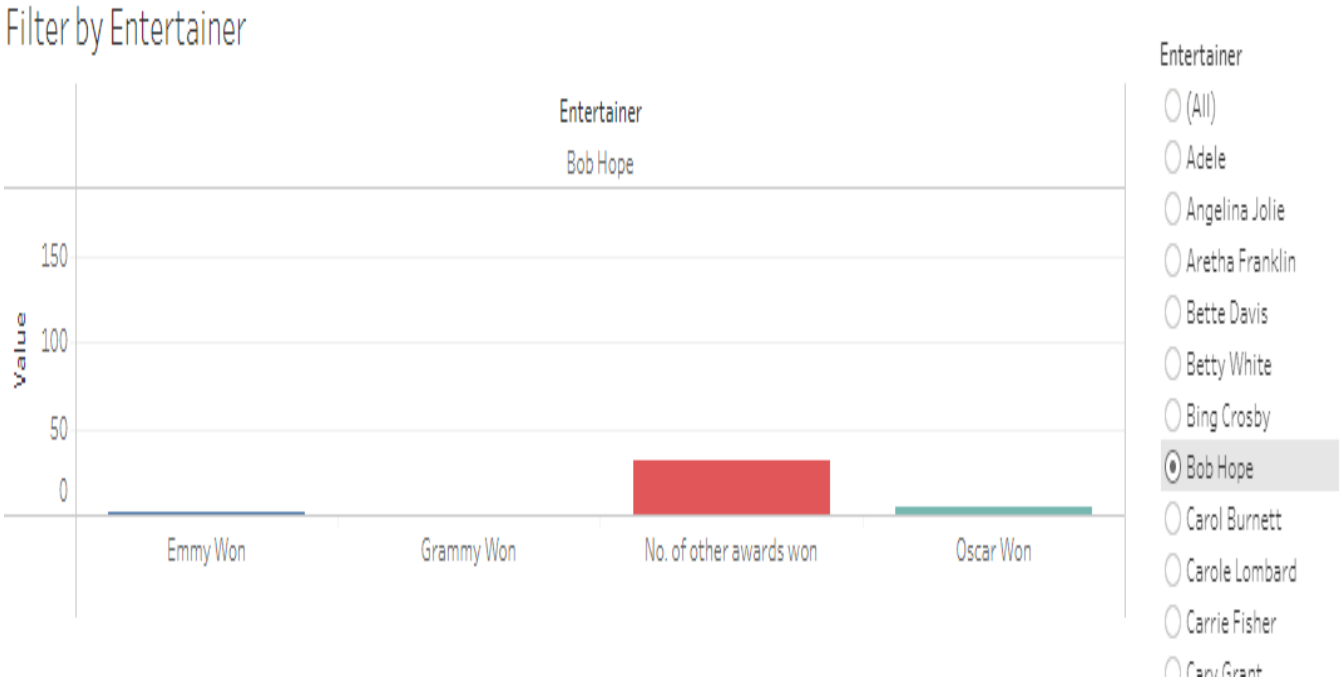
DASHBOARDS:



- Here I have merged all the sheets into Dashboard 1 including Most Awards, Most Oscars won, Classification by profession, Most Grammy, Most Emmy won. Here, in Classification by profession the orange represents singer and Blue represents Actor by profession.



This is Dashboard 2, including all Entertainers and Most Nominees



TOOLS USED:

- **Programming Languages**: Python, R
- **Data Analysis Libraries**: Pandas, NumPy, Scikit-learn
- **Visualization Tools**: Matplotlib, Seaborn, Tableau
- **Database Systems**: SQL, NoSQL databases

SOME INSIGHTS FROM THE REPORT:

- Meryl Streep is the entertainer who received highest number of awards among other entertainers
- Top 6 entertainers who received most of the awards**:**
    - Meryl Streep
    - Lady Gaga
    - Leonardo DiCaprio
    - Mariah Carey
    - Justin Timberlake
    - Will Smith
- James Dean has the highest average rating of movies among others
- Highest number of movies (55) released till date was on 1998
- Highest average rating of movies was on 1949
- Donald Sutherland acted in highest number of movies till date, which was 198 movies.
- As it is a entertainer's analysis project, based on the end user need they can consume a lot of insights from the dashboard.
- For the filtering purpose based on the end user need, in entertainers analysis page, there is a drop down filter to select the particular entertainer.
- In Movie analysis page, Included several filters like rating and year, so the end user can filter the data according to their interest.

CONCLUSION:

The Entertainer Data Analysis project aims to provide valuable insights into the entertainment industry, helping stakeholders make informed decisions. By leveraging advanced data analytics techniques, the project will uncover patterns and trends that are not immediately apparent, offering a deeper understanding of the industry's dynamics**.**

PROJECT TWO

EMPLOYER ATTRITION

INTRODUCTION

Employee attrition, also known as employee turnover, is a critical issue faced by organizations across various industries. It refers to the natural and voluntary process where employees leave the organization, leading to a reduction in the workforce. Unlike layoffs, which are often a result of organizational restructuring or economic downturns, attrition encompasses voluntary resignations, retirements, dismissals, and other forms of separation that occur over time.

Understanding and managing employee attrition is essential for several reasons. High attrition rates can lead to significant costs related to recruitment, training, and onboarding of new employees. Additionally, the departure of experienced employees often results in the loss of valuable institutional knowledge and skills, which can impact productivity and overall organizational performance. Frequent turnover can also affect the morale and engagement of remaining employees, potentially leading to further attrition.

OVERVIEW

A comprehensive analysis of employee attrition was conducted to address a persistent 15% annual turnover rate at a company established a few years ago. This high attrition rate significantly impacts productivity, team morale, and operational costs. The Employee Attrition Project aims to analyze patterns, identify key factors, and develop strategies to mitigate undesirable turnover.

OBJECTIVE

The primary objective of studying employee attrition is to identify the underlying causes and develop effective strategies to reduce turnover. By analyzing factors such as job satisfaction, work-life balance, compensation, and career development opportunities, organizations can gain insights into what drives employees to leave. This understanding enables the development of targeted interventions aimed at improving employee retention, enhancing organizational stability, and fostering a positive work culture.

Ultimately, addressing employee attrition goes beyond cost reduction; it involves sustaining a motivated and skilled workforce that propels the organization toward its objectives. By conducting

thorough analyses and adopting proactive management strategies, organizations can lessen the adverse effects of attrition, fostering a more resilient and productive work environment.

PROBLEM STATEMENT:

A Company which was established a few years back is facing around a 15% attrition rate for a couple of years. And it's majorly affecting the company in many aspects. In order to understand why employees are leaving the company and reduce the attrition rate, we proposed conducting a HR analytics study. This data-driven approach will enable the company to make informed decisions, potentially reducing the attrition rate to below 10% within the next fiscal year and saving financial losses for the company and building a dashboard which can help the organization in making data-driven decisions.

DATASET INFORMATION:

The dataset will be used to:

- Identify patterns and trends in employee attrition.
- Analyze demographic and job-related factors that contribute to higher attrition rates.
- Develop predictive models to identify employees at risk of leaving.
- Create visualizations and dashboards to support data-driven decision-making.
- Propose targeted interventions to improve employee retention.
- Below is an example of the type of information that should be included in the dataset:

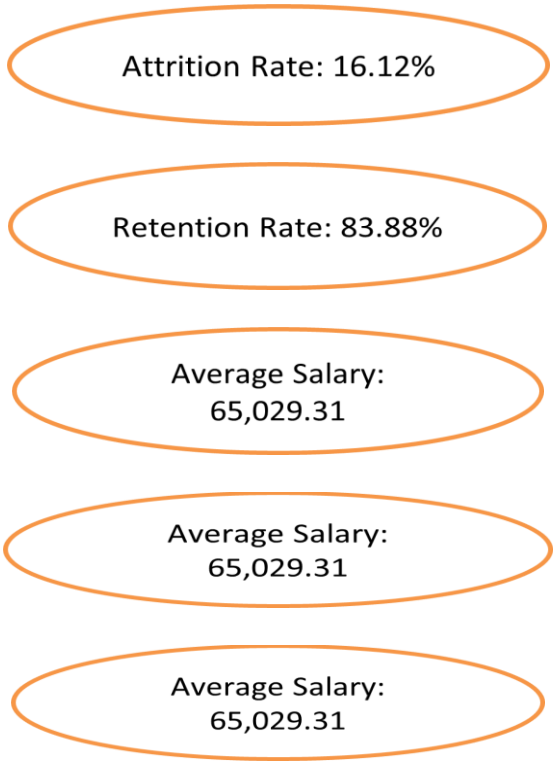
    - Employee information
    - Employee details
    - Work Environment
    - Engagement and satisfaction
    - Benefits and perks
    - Reason for leaving
    - Demographic information

TOOLS USED:

- **Python –** For Data cleaning and Exploratory Data Analysis (EDA)
- **Pandas** - For Data manipulation and analysis
- **Matplotlib and Seaborn** - For Data Visualization

- **Power Bi** - For Data visualization using Dashboards
- **Scikit-Learn** - For Feature Analysis and Predictive Analysis

KEY PERFORMANCE INDICES (KPI'S):



- Key Performance Indicators (KPI's) are critical metrics that help in evaluating the effectiveness and overall health of the organization.
- It provides insights into various aspects such as attrition and retention rate, salary, performance, and employee satisfaction.
- High attrition rate in certain departments while average ratings on job satisfaction and work-life balance can guide improvements in workplace policies and practices.

DATA CLEANING:

Data Cleaning process involves fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset, to prepare data for analysis.

- **Dropping Unnecessary Columns:** Three columns have been identified where the values remain constant across all entries, rendering them irrelevant for analysis. These columns will therefore be dropped.

- **Handling Missing Values**: Identifying and dealing with missing data points in the dataset, which could involve imputation techniques or deciding to drop incomplete records based on the impact on analysis.
- **Standardizing Formats**: Ensuring consistency in data formats across different fields, such as dates or categorical variables, to facilitate accurate analysis and modeling.
- **Removing Duplicates Entries**: Identifying and removing duplicate entries that can skew analysis results and create biases in attrition predictions.

KEY FINDINGS:

1. **Exit Interviews:** Exit interview data highlights common reasons for leaving the company, such as employees find better job offers elsewhere, have a lack of career advancement, poor management practices, and work-related stress.

2. **Predictive Factors:** Predictive modeling identified key factors contributing to attrition, including low job satisfaction, limited career advancement, work-life imbalance, and low compensation. These factors can be used to identify at-risk employees.

3. **Demographic Insights:** Younger employees (18-22 years) and single individuals showed significantly higher attrition rates. Male employees exhibited a slightly higher tendency to leave (16.67%) compared to females (15.31%).
4. **Impact of Wok environment:** Work environment factors, such as physical workspace, company culture, and team dynamics, also play a significant role in employee retention. Poor work environments contribute to higher attrition rates.
5. **Salary and financial Incentives:** Monthly income emerged as the most influential factor (importance score: 0.097). Employees with lower stock option levels and minimal salary hikes were more likely to leave. Surprisingly, higher salary hikes (21-25%) correlated with increased attrition, suggesting factors beyond compensation influence turnover.
6. **Career Progression:** Employees at lower job levels and those with fewer years since their last promotion showed higher attrition rates. Lack of training opportunities significantly increased turnover risk.
7. **Commute and Travel:** Employees with longer commutes (especially around 20km) had higher chance of attrition rates. Frequent business travellers were more likely to leave (23.78% attrition) compared to non-travelers (9.02).

PREDICTIVE ANALYSIS:

- **Increase in Monthly Income:** Increased 15% in monthly income for employees, 10% increased in percent salary hike, and increased stock option level by 1. These changes predicted a 6.47% decrease in the attrition rate, reducing it from 16.15% to 15.09% identified as high risk to improve retention.
- **Job Satisfaction Improvements:** Enhance work-life balance, provide career development opportunities, and recognize employee achievements.
- **Competitive Compensation:** Offer competitive salaries and benefits to retain top talent.
- **Targeted Interventions:** Focus retention efforts on high-risk groups, such as new hires and employees in high-turnover departments.
- **Enhanced Onboarding:** Improve the onboarding process to ensure new hires are well-integrated and supported.

LIMITATIONS:

- **Data Quality Issues**: The accuracy and reliability of predictions heavily depend on the quality of the data used. Incomplete, outdated, or biased data can lead to inaccurate insights and predictions.
- **Limited Scope**: The analysis focuses primarily on internal company data and may not account for external factors such as industry trends, economic conditions, or regional job markets.
- **Complexity of Factors**: Employee attrition is influenced by a wide range of factors such as job satisfaction, work-life balance, compensation, career development, and organizational culture. Capturing all relevant factors and their interactions can be challenging.
- **Changing Dynamics**: Workforce dynamics can change rapidly due to internal and external factors (e.g., economic conditions, industry trends, company policies). Models trained on historical data may not always capture these changes effectively.
- **Privacy and Ethical Concerns**: Employee attrition projects often involve sensitive personal information. Ensuring data privacy and ethical use of data is crucial to maintain trust and comply with regulations such as GDPR or CCPA.
- **Overfitting**: Machine learning models may overfit to historical data, making them less effective at predicting future trends accurately. Regular model validation and tuning are essential to mitigate this risk.

CONCLUSION:

Predictive analysis of employee attrition provides valuable insights for reducing turnover and improving employee retention. By identifying key factors such as job satisfaction, compensation, and career development, organizations can implement targeted strategies to address these issues. Despite limitations like data quality, bias, and dynamic workforce changes, predictive models can significantly enhance decision-making. Continuous refinement of models and ethical data use are essential for creating a resilient and productive work environment.