Talent Guardian: Preventing Employee Attrition with ETL, DW, and OLAP

Frederico Cardoso, Naomi Esposito, and Eduardo Neves

Department of Informatics Engineering, University of Coimbra, Portugal
Data Analytics Technology, Master's in Data Science and Engineering
December 20, 2023

Abstract. Employee attrition poses a significant challenge for organizations, impacting talent retention and incurring associated costs. This work introduces an AI tool that aims to predict employees leaving a company. Utilizing the IBM HR Analytics dataset, the project used data warehousing, as well as machine learning, to save valuable information and make informed predictions. The framework offers a proactive approach to workers management, providing actionable insights for organizations to optimize workforce strategies.

Keywords: attrition · employees · data warehouse · artificial intelligence

1 Introduction

1.1 Context

The world is in a constant state of change, presenting organizations with a multitude of challenges to maintain its pace. Nevertheless, employees still stand as a fundamental pillar of a company's resilience and adaptability, even with the rise of artificial intelligence technologies. As such, **employee attrition** stands as a substantial challenge.

Ensuring that employees operate within a dynamic and conductive environment, with the essential conditions is more important than it sounds. A happy, satisfied and focused workforce inherently becomes more productive and committed.

1.2 Relevance

Bill Gates once remarked, "You take away our top 20 employees and we [Microsoft] become a unimportant company". This sentiment holds true even beyond top-tier employees. The unexpected departure of valuable team members can, not only disrupt the workflow and growth of organizations, but also create greater losses than one would anticipate.

In fact, the average cost of turnover per employee can amount to thousands of dollars. Some studies suggest that **every time a business replaces a salaried employee, it costs 6 to 9 months' of their average salary** [1]. Acknowledging the significance of this challenge, we also identified the potential to develop a practical solution, focused on mitigating employee attrition proactively.

2 Motivation

The intricacies of employee's status, details and satisfaction, recognized as a pivotal factor in prediction attrition, present a real-world challenge that aligns seamlessly with the core principals of the Data Analytics Technology course. As organizations grapple with talent retention and the aforementioned associated costs of employee departure, we discern a significant opportunity.

Through harnessing critical employee characteristics, we aim to construct and provide a **business solution designed on forecasting the likelihood of an employee departing in a near future**. This solution, centered on Artificial Intelligence (AI) and Data Warehousing (DW), is envisioned to be a revolutionary asset for companies seeking proactive measures in talent management.

2.1 Methodology

This section outlines the expected workflow, designed to guide the progression of this work, commencing with the acquisition of relevant data, and culminating in the deployment of a predictive business solution. Each step is structured not only to ensure coherence and efficiency thorough the process, but also handle the integration of different technologies.

The key phases include data acquisition, transformation, loading, OLAP-based analysis and a comprehensive description of major challenges and results. Subsequent chapters will unfold deeper exploration in each of these phases, shedding light into the methodologies employed and the outcomes achieved.

ETL (Extract, Transform, Load) is the process of combining data from multiple sources into a large, central repository called a data warehouse. It serves as a cornerstone in our methodology. ETL uses a set of business rules to clean and organize raw data and prepare it for storage, data analytics, and Machine Learning (ML) [2]. ETL also automates repeatable data processing tasks for efficient analysis [2]. This automated migration process guarantees data accuracy, streamlines subsequent updates, laying the groundwork for effective data analysis and predictive modeling.

OLAP (Online Analytical Processing) employs a multidimensional strategy for organizing and analyzing data, delivering dynamic and interactive insights. OLAP tools facilitate efficient processing of a growing volume of digital information, enabling faster decision-making, non-technical user support, and integrated data visualization [3]. Within our solution, OLAP serves as a vital tool for navigating through retrieved data. Tailored for handling complex queries and aggregations, OLAP databases enable swift exploration of vast datasets, deepening our understanding of trends and patterns. Our reliance in OLAP is particularly evident in tasks such as visualizing departmental distributions, identifying at-risk employees, and extracting actionable insights. By applying OLAP principles, we empower decision-makers to interact dynamically with data, fostering informed decision-making and enhancing the overall utility of our solution.

2.2 Tools

This section introduces the utilized tools in this project.

Python was the chosen programming language to the extraction and transformation phases. Its versatility and extensive libraries facilitated effective data manipulation, feature engineering and forecasting with well-know ML models.

PostgreSQL acted as the relational database management system, ensuring efficient data storage and retrieval. Its scalability, reliability, and support for complex queries were essential for managing the structured data generated [4]. Also, the ***pgAdmin*** application was chosen, as it provides management tools for the PostgreSQL relational database management system [5].

Pentaho played a vital role in data extraction, transformation, and loading processes, smoothly integrating with the PostgreSQL database system [6]. This open-source ETL tool acted as the primary data retrieval tool.

Tableau emerged as the key visualization tool, providing dynamic and interactive dashboards. Its capabilities facilitated comprehensive data exploration, aiding in the interpretation of results, and offering insights to decision-makers. Besides that, it also features a connector to PostgreSQL database systems [7].

3 Extraction

3.1 Dataset

For addressing the challenge posed by employee attrition, we used the **IBM HR Analytics Employee Attrition & Performance** dataset. This fictional dataset, curated by IBM data scientists, serves as a valuable resource to unravel the factors influencing employee attrition. Sourced from the *Kaggle* [8] platform, it was designed to explore important questions such as ”show me a breakdown of distance from home by job role and attrition” or ”compare average monthly income by education and attrition”.

This dataset's applicability aligns with our project objectives, as we have a similar goal to the one for which it was created. While we retain binary classification for attrition, our unique contribution lies in extending this classification to a percentage prediction.

The dataset encompasses 1470 entries, featuring a diverse set of 35 of both categorical and numerical attributes, that can be classified as:

- **Personal details**: Age; DistanceFromHome; Education; EducationField; Gender; MaritalStatus; NumCompaniesWorked; Over18; TotalWorkingYears.
- **Contract details**: BusinessTravel; DailyRate; HourlyRate; MonthlyRate; MonthlyIncome; PercentSalaryHike; StandardHours; StockOptionLevel.

- **Job details**: Department; JobLevel; JobRole.
- **Employee mental status**: EnvironmentSatisfaction; JobSatisfaction; RelationshipSatisfaction; WorkLifeBalance.
- **Employee identification**: EmployeeCount; EmployeeNumber.
- **Employee job status**: JobInvolvement; OverTime; PerformanceRating; TrainingTimesLastYear; YearsAtCompany; YearsInCurrentRole; YearsSinceLastPromotion; YearsWithCurrManager.
- **Attrition**: The binary feature that will be used in the prediction process.

3.2 Relevant Features

With this classification of attributes, we're already gaining insights into the relevant dimensions for the later stages of the project. Notably, certain features, such as "Over18", "EmployeeCount" and "StandardHours" seem completely irrelevant at that point. Additionally, it becomes apparent that the mental status characteristics may duplicate the purpose of the "Attrition" attribute, potentially offering limited value to the ML tasks ahead.

As we proceed, we will evaluate and potentially exclude less informative features to ensure a focused and efficient approach to both the predictive modeling phase and to the data warehousing design and organizational process.

4 Transformation

4.1 Data Preprocessing

The initial step in our data transformation process involved examining the dataset for any missing values. Fortunately, every record within the dataset contains valid entries, eliminating the need for common strategies such as deletion or imputation (e.g., using metrics like the average or the most common value). Additionally, each attribute is clearly defined in terms of its data type—whether *integer*, *float*, *string*, or other categories. As a result, there was no necessity for additional processing in this regard.

One crucial consideration is that this dataset is highly imbalanced, having 237 positive values for attrition out of a total of 1470 records (16.1%). This implies that, when training ML models, the class distribution may lead to a bias where the model becomes proficient at predicting the majority class, but struggles with the minority class - the very class of interest [**?**]. Therefore, we will employ resampling techniques and specialized algorithms designed for handling imbalanced datasets. Through these measures, we aim to develop a model that effectively generalizes across both classes.

4.2 Feature Engineering

The last chapter emphasized the exclusion of the features "Over18", "EmployeeCount" and "StandardHours" due to the lack of their informative value. Similarly, features directly related to the employee satisfaction, are automatically discarded for its usage in the ML models. Here are the remaining ones:

- **Personal details**: Age; DistanceFromHome; Education; EducationField; Gender; MaritalStatus; NumCompaniesWorked; TotalWorkingYears.
- **Contract details**: BusinessTravel; MonthlyIncome; PercentSalaryHike; StockOptionLevel.
- **Job details**: Department; JobLevel; JobRole.
- **Employee job status**: JobInvolvement; OverTime; PerformanceRating; TrainingTimesLastYear; YearsAtCompany; YearsInCurrentRole; YearsSinceLastPromotion; YearsWithCurrManager.
- **Attrition**: The binary feature that will be used in the prediction process.

Additionally, two new features were introduced to enhance the predictive capabilities of the model:

- **AttritionPercentage**: This attribute serves as the primary the goal of our solution, indicating the probability of an employee leaving the company.
- **ExpectedLeavingCost**: Calculated based on the literature review, this feature estimates the expected financial cost to the company in the event of an employee departing.

The incorporation of these additional features enhance the precision and effectiveness of our dataset, aligning it with the project's objectives.

4.3 Calculating Attrition Probability

In this segment, we delve into a crucial aspect of our solution – predicting the probability of an employee departing. To accomplish this, we employed four distinct models: Logistic Regression, Random Forest, Support Vector Machines (SVM), and XGBoost. Each model possesses its unique strengths and weaknesses. Therefore, we conducted comprehensive tests across all models to compare results and identify the most suitable one.

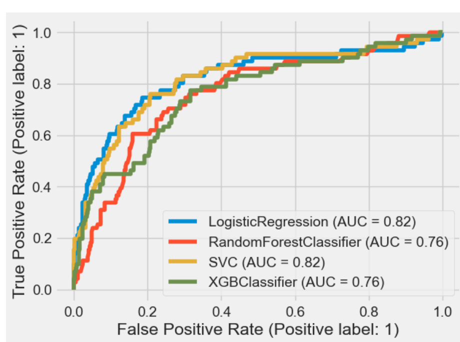

Fig. 1: Testing on which model would be the most efficient for prediction.

Following careful implementation and evaluation (refer to figure 1), the selected model demonstrates superior performance and precision in estimating attrition likelihood. This model will consistently be employed for calculating and presenting this probability with each subsequent new entry in the database.

It's important to note that among the selected models, Logistic Regression, Random Forest, and XGBoost are equipped to predict probabilities, including the probability of attrition. Since these models provide probability estimates for each class, these probabilities can be interpreted as the likelihood of an observation belonging to a specific class. In binary classification problems like attrition prediction, it's a conventional practice to utilize the predicted probability of the positive class. However, Support Vector Machines (SVM), in its basic form, is a classification algorithm that outputs class labels and doesn't inherently offer probability estimates. While there exists a variant called SVM with Probability Estimates (SVM-Prob) capable of estimating probabilities, it is less commonly used for probability estimation compared to the other mentioned models.

While the chosen model exhibits the best performance, it's essential to acknowledge any limitations and consider avenues for future improvement. Subsequent iterations of this project may involve fine-tuning the model parameters, exploring advanced machine learning techniques, and incorporating additional features to enhance the accuracy and applicability of attrition probability predictions.

4.4 Calculating Expected Cost of Replacement

The Expected Cost of Replacement (ECR) is a metric that we created that estimates the financial impact associated with employee attrition. It is calculated via the annual salary, which translates into 12 times the Monthly Income (MI), and scaled by 68%:

$$ECR = 12 \times MI \times 68\%$$

This scaling factor represents the mean cost of replacement, calculated through the mean value of multiple sources [1] [9] [10] [11] [12] [13] [14].

5 Loading

The loading phase, as explained before, refers to the process of populating a data warehouse with data from singular or multiple sources.

A **star schema** is a type of data warehouse schema where a central fact table is connected to multiple dimension tables through foreign key relationships. The fact table contains quantitative data (measurements or metrics), and the dimension tables contain descriptive information related to the data in the fact table. The structure resembles a star when visualized, with the fact table at the center and dimension tables radiating outward [15].

The importance of a star schema lies in its simplicity, ease of understanding, and query performance. It provides a straightforward way to model relationships between data, making it intuitive for users to navigate and query. The star schema's design optimizes query performance by reducing the number of joins needed to retrieve data, leading to faster and more efficient analytics. For that reason, it suits perfectly our project, to facilitate retrieval of information from our DW in a simple and efficient way.

5.1 Star Schema

Regarding the objective of our solution, the star schema should be structured to effectively address the questions posed in the initial deliverables, as outlined in the Practical Applications 7 chapter. A key aspect is to systematically organize the star schema, distinguishing between internal and external factors affecting the company. Our proposed star schema can be observed in figure 2.

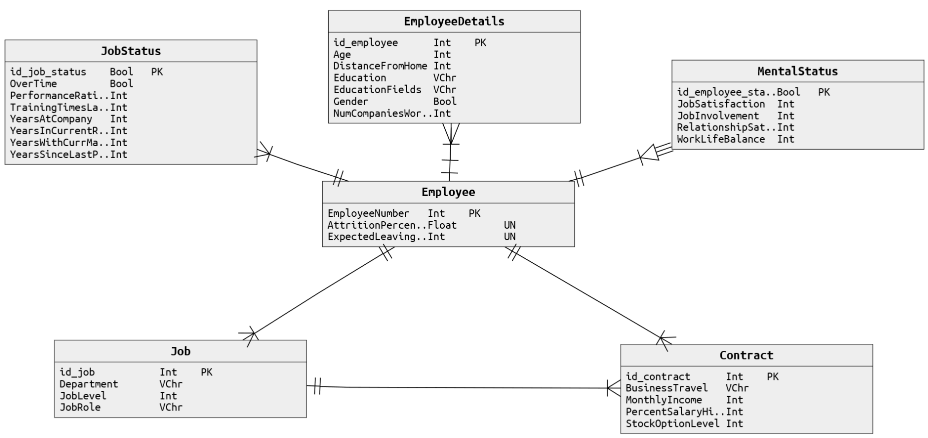

Fig. 2: The proposed star schema, developed in `https://onda.dei.uc.pt/v4/`. Although the application supports the usage and generates foreign keys, they are not illustrated in the schema itself, when relationships are visible.

5.2 Dimensions

In the fact table, each dimension is associated with surrogate keys (four keys), alongside the primary identifier, "EmployeeNumber.".

Employee (Fact Dimension) As the focus of the work the employee is placed in the middle as the fact table. Besides the EmployeeNumber got from the data as a primary key, it features the two attributes of study. This provides easier access for the OLAP while improving speed in getting these values of interest.

Employee Details This dimension server the purpose of gathering personal details of the workers. This is separated from the fact table to enhance navigation an queries.

Job Here lie the information about each individual job an employee might be assigned to. It is related to the contract as each job has a contract.

Contract The dimension gathers all the fields related to information that might be in the contract of a worker. The relation with the Job dimensions is one to many since each contract is unique, but the job might be the same.

Mental Status The dimension includes some fields that employees might feel in the times at the company.

Job Status This dimension is related to the current status of the employee at the company. It is separated from the Job dimension since the latter describes more technically the position while the former relates to the situation on the company.

5.3 Relationships

The relationships, on the perspective of the Employee table, are one-to-one as one employee can have only one of each dimension fields. The reverse is counted as one-to-many since, even unlikely, the same fields for each of all dimensions might fit two different workers.

In this schema we considered the we consider the MentalStatus weak as the relationship is not straightforward to the fact table. This is because it contains attributes related to satisfaction of the employees and is based on surveys and feedback.

5.4 Indexing

Indexing, in the context of databases, involves creating a data structure that enhances the speed of data retrieval operations on a database table. It works by providing a quick reference to the location of data within the table, optimizing query performance.

Additionally, in our setup, primary keys and foreign keys in each dimension are routinely indexed. However, a deliberate choice was made to extend indexing to the two features in the facts dimension, *i.e.*, **AttritionPercentage** and **ExpectedLeavingCost**. This decision is rooted in their frequent use in data retrieval operations, aiming to optimize query performance and enhance the overall efficiency of the system.

5.5 Evaluation

With the Pentaho application, we could gather insights about some performance metrics to evaluate the ETL phase, as seen in the table below.

Metrics of DW		
size of data on load	**time 1^{st}**	**time for update**
31 cols	$1487ms$	$946ms$

Table 1: Metrics of the data warehouse loading. **A**: size of table before the population of dimensions; **B**: time of populating the data warehouse on the first time; **C**: time of updating values on each upload.

The performance metrics suggest that the ETL process, as implemented in the Pentaho application, demonstrates reasonably efficient loading times for both the initial load and subsequent updates. The number of columns indicates a moderately complex data structure. However, our dataset had a small initial size, comprising only 1470 entries. Further analysis and comparison with more data and/or predefined benchmarks or requirements can provide a more comprehensive assessment of the ETL process's performance.

5.6 Final Considerations

Following the first load, the update strategy for dimensions is an important step to choose. For this work we consider a type 4 strategy involves creating a separate mini-dimension table to store historical changes. The importance of storing historical data lies on the relevance to the model. Even if the algorithm predicts with real-time data, this data can be gathered and then used for training and refinement.

For the fact table the solution might be incremental load as only new or changed data since the last update is loaded. This allows for more versatility of the data warehouse since it can adjust easily for bigger companies with bigger employees as a full load or similar will deal with too much information.

The frequency of updates is not a pressing topic since the the data is not expected to change rapidly. The fact table might change when there is recruitment or firings at the company. The dimensions might change when a employee receives a new contract or gets married. All those situations are spread evenly in time and some, like recruitment, normally change at the same time. There are also some fields that change yearly.

6 Online Analytical Processing

In this section, we explain how our system facilitates a seamless experience for end-users, particularly from a business perspective. The key operations performed in OLAP—querying and visualizing data—are orchestrated through a streamlined process that combines Pentaho, pgAdmin, and Tableau.

6.1 Querying and Visualizing the Data

Modelling and Pentaho Integration. Pentaho serves as a fundamental tool, as it is with it that we model the requirements into a structured SQL query.

Export to pgAdmin. The SQL query, crafted and refined within Pentaho, is then exported to pgAdmin, that can process those same SQL queries to retrieve the relevant data.

Tableau Integration for Visualization. After the querying operation, exported data is then seamlessly integrated into Tableau, enabling users to explore insights through a variety of dashboards and visualizations.

Navigating Dashboards in Tableau. End-users are then presented with an array of dashboards and visualizations within Tableau, offering a rich and interactive experience. Figures 5, 7, 9, 11, and **??** showcase examples of the visualizations available. These dynamic displays empower users to delve into different dimensions of the data, gaining valuable insights into various aspects of employee attrition.

6.2 Business Point of View

From a business perspective, our solution holds significant importance for companies seeking proactive measures in talent management and attrition mitigation. The value proposition extends across several key aspects:

Predictive Insights for Strategic Decision-Making: Our solution leverages advanced Machine Learning models to predict the likelihood of employee attrition. This empowers decision-makers with valuable insights for strategic workforce planning and talent retention strategies.

Visualization of Key Metrics and Trends: Through the integration of Tableau for data visualization, companies can access dynamic and interactive dashboards. These dashboards present key metrics and trends related to employee attrition, enabling a comprehensive understanding of workforce dynamics.

Efficient Data Exploration and Analysis: The OLAP capabilities incorporated in our solution facilitate efficient exploration and analysis of vast datasets. Decision-makers can navigate through different dimensions, analyze trends, and gain actionable insights to inform data-driven decisions.

Customized Recommendations for Talent Management: As part of future work, our solution envisions offering recommendations to companies based on predictive analytics. These recommendations could include tailored strategies to reduce attrition, optimize employee contracts, and enhance overall employee satisfaction.

User-Friendly Interface for Accessibility: The user interface is designed to be user-friendly, ensuring accessibility for decision-makers across various departments. Whether in HR, management, or other relevant roles, users can interact with the system effortlessly to gain insights into employee attrition factors.

Cost Estimation for Informed Decision-Making: The calculation of the Expected Cost of Replacement (ECR) provides companies with an estimate of the financial impact associated with employee attrition. This information is crucial for budgeting and resource allocation, offering a proactive approach to managing workforce transitions.

Scalability and Adaptability: The solution is built on scalable technologies such as PostgreSQL and Pentaho, ensuring adaptability to the evolving needs and scale of the organization. As companies grow and evolve, the solution can accommodate larger datasets and additional features.

In essence, our solution goes beyond traditional HR analytics by providing a holistic and a much more intuitive and visual approach to employee attrition, as we see in the figure below 3. By integrating predictive modeling, data visualization, and actionable insights, companies can proactively address attrition challenges, foster a positive work environment, and strategically manage their workforce for sustained success.

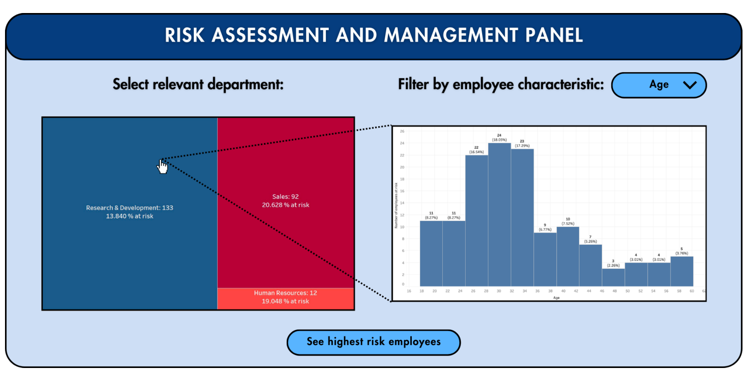

Fig. 3: An illustrative example of a useful dashboard for discovering Attrition levels by Department.

7 Practical Applications

7.1 What is the average Performance Rating by Job Role within each Department?

```
SELECT
    e.Department,
    e.JobRole,
    AVG(js.PerformanceRating) AS Avg_Performance_Rating
FROM
    Employee e
JOIN
    Job_Status js ON e.EmployeeNumber = js.EmployeeNumber
GROUP BY
    e.Department, e.JobRole;
```

Fig. 4: SQL query to retrieve the necessary data to answer question 1.

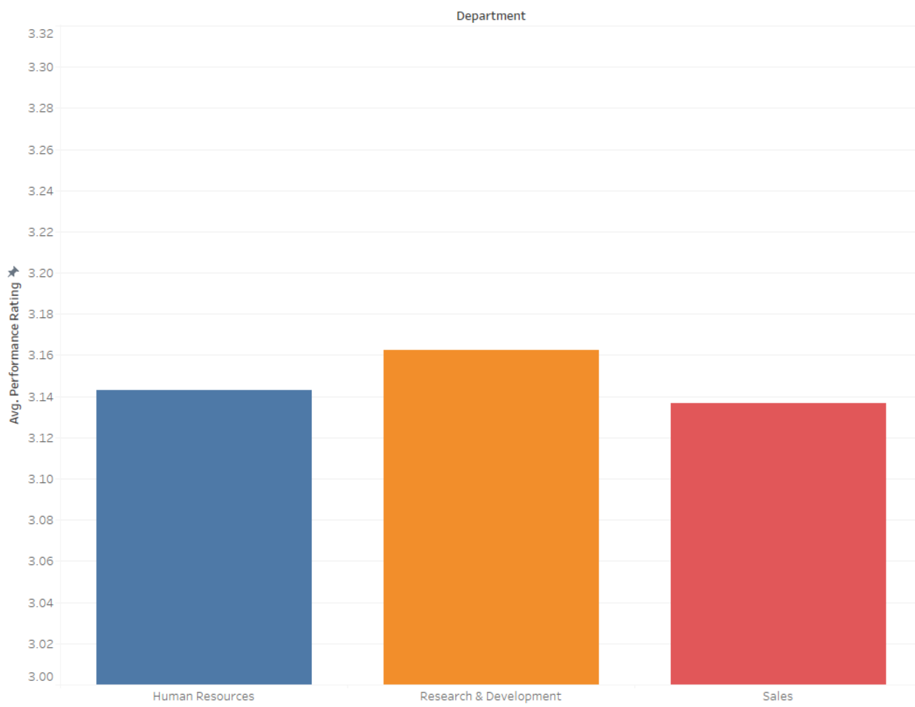

Fig. 5: Tableau visualisation of data retrieved from query above.

7.2 What is the average Age and Years in Current Role, by Job Role?

```sql
SELECT
    JobRole,
    AVG(Age) AS Avg_Age,
    AVG(YearsInCurrentRole) AS Avg_YearsInCurrentRole
FROM
    Employee
GROUP BY
    JobRole;
```

Fig. 6: SQL query to retrieve the necessary data to answer question 2.

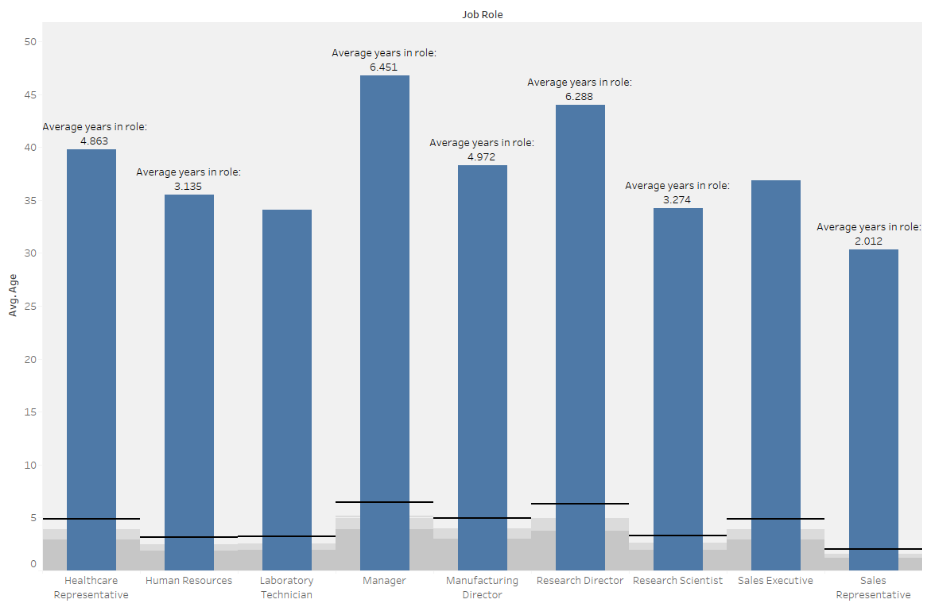

Fig. 7: Tableau visualisation of data retrieved from query above.

7.3 How many Employees at Risk are in each Department?

```sql
SELECT
    Job.Department,
    COUNT(Employee.EmployeeNumber) AS TotalEmployees,
    SUM(CASE WHEN Employee.Attrition > 0 THEN 1 ELSE 0 END) AS AttritionEmployees,
    AVG(Employee.Attrition) AS AvgAttritionPercentage
FROM
    Job
JOIN
    Employee ON Job.id_job = Employee.id_job
GROUP BY
    Job.Department;
```

Fig. 8: SQL query to retrieve the necessary data to answer question 3.

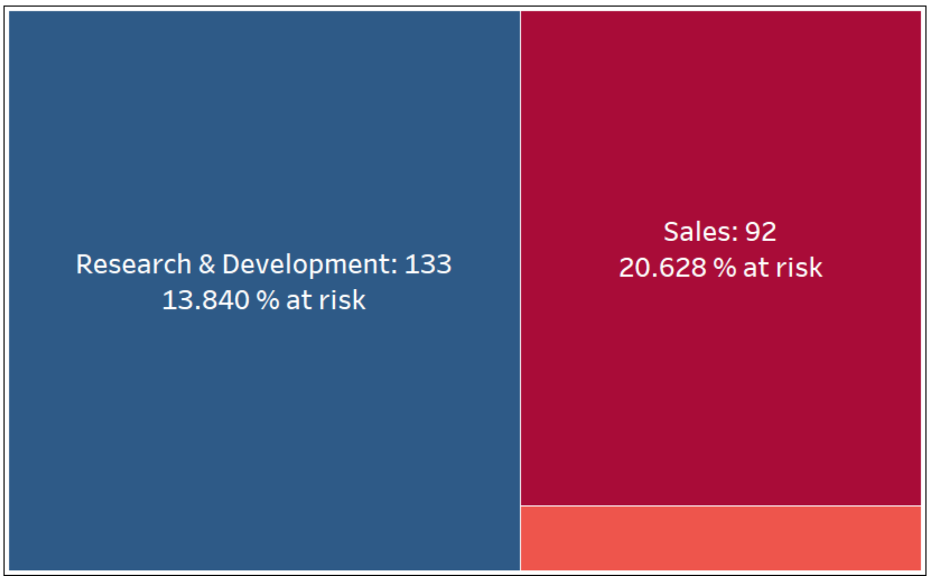

Fig. 9: Tableau visualisation of data retrieved from query above.

7.4 How many Years in the Company and in the same Job Role are the Employees in each Department?

```sql
SELECT
    Job.Department,
    AVG(JobStatus.YearsAtCompany) AS AvgYearsAtCompany,
    AVG(JobStatus.YearsInCurrentRole) AS AvgYearsInCurrentRole,
    AVG(JobStatus.YearsWithCurrManager) AS AvgYearsWithCurrManager
FROM
    Job
JOIN
    JobStatus ON Job.id_job = JobStatus.id_job_status
GROUP BY
    Job.Department;
```

Fig. 10: SQL query to retrieve the necessary data to answer question 4.

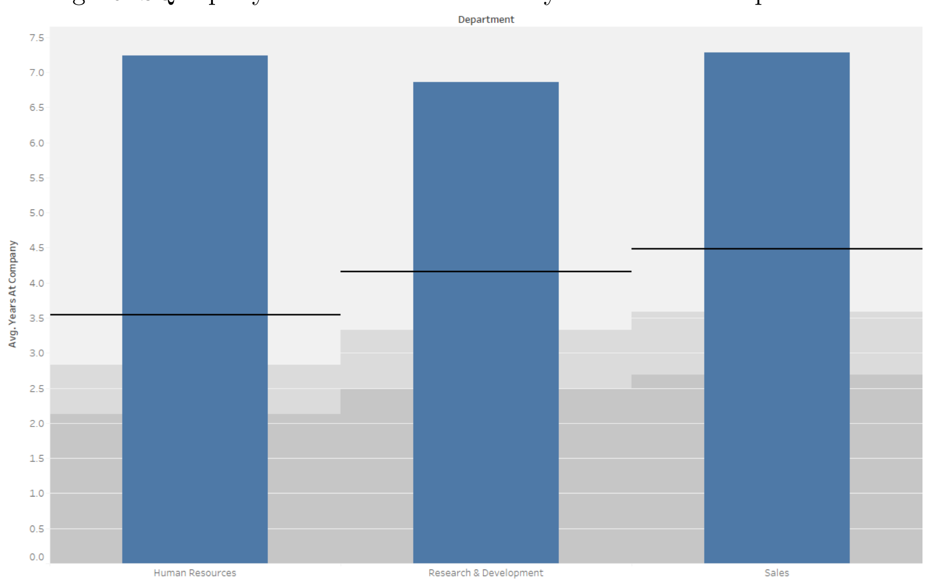

Fig. 11: Tableau visualisation of data retrieved from query above.

8 Conclusion

Our project aims to tackle the real issue of employee attrition using advanced data analytics and machine learning. Leveraging a diverse dataset, we carefully engineered features and created new metrics to enhance our predictive capabilities. We also extended the traditional binary classification to predict attrition percentages, as it benefits better an organization.

Machine learning models, including Logistic Regression, Random Forest, SVM, and XGBoost, were employed for accurate attrition probability predictions. The Logistic Regression model demonstrated superior performance, becoming the foundation of our predictive analytics.

Our solution extends beyond model implementation, incorporating a star schema data warehouse plan for efficient data retrieval and analysis through OLAP. Tools like Pentaho, PostgreSQL, and Tableau ensured a seamless integration to translate the raw data to a user-friendly interface for decision-makers.

From a business perspective, our solution empowers strategic decision-making with predictive insights, key metric visualizations, and tailored talent management recommendations. The Expected Cost of Replacement adds a financial dimension, aiding informed decisions.

Looking ahead, opportunities for fine-tuning models, exploring advanced techniques, temporal analysis, and feedback loop implementation are identified. Our scalable and adaptable solution positions itself as a valuable asset for organizations navigating talent management challenges, fostering a resilient and engaged workforce.

9 Future Work

As we conclude this project, several avenues for future work emerge, offering opportunities to enhance the robustness and applicability of our solution to predict employee attrition. These could be:

- Fine-Tuning model parameters
- Advanced ML techniques
- Temporal analysis
- Feedback loop implementation
- Advanced feature engineering
- External data integration
- Employee engagement strategies
- Ethical considerations
- User interface and deployment
- Validation on external datasets

One noteworthy avenue for future exploration involves extending our solution to provide actionable recommendations to the company. Although not initially included in our proposed steps, we presented in our pitch the idea of the system offering recommendations to the company on ways to reduce attrition probabilities (see figure 12).

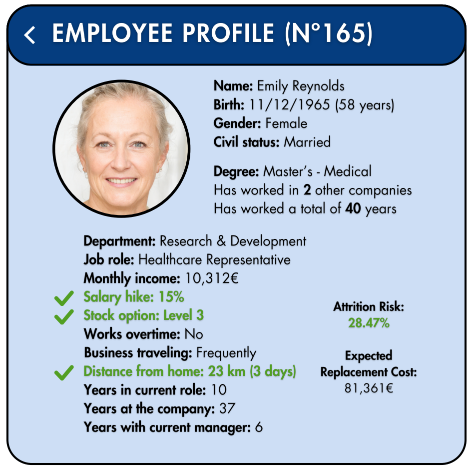

Fig. 12: A suggested refinement for our current solution entails the system delivering recommended actions for the company to take.

References

1. Charaba, C. (2023, August 3). Employee retention: The real cost of losing an employee. *People Keep*. https://www.peoplekeep.com/blog/employee-retention-the-real-cost-of-losing-an-employee.
2. What is ETL (Extract Transform Load)? (n.d.). *Amazon AWS*. https://aws.amazon.com/what-is/etl/
3. What is OLAP ((Online Analytical Processing)? (n.d.). *Amazon AWS*. https://aws.amazon.com/what-is/olap/
4. About. (n.d.). *PostgreSQL*. https://www.postgresql.org/about/
5. F.A.Q.. (n.d.). *pgAdmin*. https://www.pgadmin.org/faq/#1
6. Use PostgreSQL as your repository database. (2023, November 22). *Pentaho*. https://help.hitachivantara.com/Documentation/Pentaho/Data_Integration_and_Analytics/8.3/Setup/Use_PostgreSQL_as_your_repository_database_(Manual_installation)
7. PostgreSQL Connector. (n.d.). *Tableau*. https://help.tableau.com/current/pro/desktop/en-us/examples_postgresql.htm
8. IBM HR Analytics Employee Attrition & Performance. (2016). *Kaggle*. https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset.
9. Andrea Dumont. (n.d.). What is the Cost of Employee Turnover for your Business? *Enboarder*. https://enboarder.com/blog/cost-of-employee-turnover/
10. What is the true cost of replacing an Employee? (2022, January 21). *Centric HR*. https://www.centrichr.co.uk/what-is-the-true-cost-of-replacing-an-employee/
11. The Real Cost of Employee Turnover (And How to Prevent It). (2023, November 26). *Jobvite*. https://www.jobvite.com/blog/cost-of-employee-turnover/
12. How much does it cost to resign an employee?. (n.d.). *Amelio*. https://amelio.co/cost-hire-employee/

13. Fluckiger, B. & Ruddy, C. (n.d.). Employee Replacement Costs. *Eddy*. https://eddy.com/hr-encyclopedia/employee-replacement-costs/
14. How Much Does It Cost to Lose an Employee? (n.d.) *Skillworld*. https://resources.skillwork.com/cost-of-hiring-new-employee-vs-retaining
15. Krippendorf, M. and Il-Yeol Song, Database and Expert Systems Applications. 8th International Conference, DEXA '97. Proceedings, "The translation of star schema into entity-relationship diagrams"