REPORT

Team Members:

Saumya Kumar	2401010432
BulBul Agarwalla	2401010131
Jatin Verma	2401010200
Yachna	2401010509
Supreet	2401010465
Aditya Ranjan	2401010035

Sector :Corporate
Institute: Rishihood University
Mentor : Satyaki Das

Executive Summary

Overview: This report presents a comprehensive analysis of employee attrition utilizing a dataset of 10,025 employee records. The primary objective is to identify key drivers of workforce turnover and provide data-backed recommendations to improve retention rates.

Problem

Corporate attrition represents a significant cost center, estimated at 1.5x to 2x an employee's annual salary. The organization currently faces an attrition rate of approximately 18.15%, with notable spikes in specific departments such as Sales and Research & Development. Management lacks visibility into the specific correlations between employee demographics, job satisfaction, and their decision to leave.

Approach

Our team employed a structured analytics framework using the Kaggle Employee Attrition Prediction dataset. The methodology involved:

- Data Cleaning: Handling missing values, duplicate rows, not optimize attributes.
- Exploratory Data Analysis was conducted to identify patterns and relationships across key variables such as age, monthly income, tenure, job satisfaction, department, and overtime status. Various charts and summary statistics were used to compare attrition trends across employee groups.
- Dashboard: An interactive dashboard was developed using Google Sheets to monitor key performance indicators including attrition rate, average income, and employee tenure. The dashboard includes dynamic filters for department, job role, gender, and overtime status to support customized analysis.
- Analytical Insights:Analytical insights were derived by examining correlations and comparisons among demographic and job-related variables. High-risk employee segments were identified based on age group, income level, tenure, department, and workload patterns.

Key Insights

- Overtime is a Critical Factor: Employees working overtime show relatively higher attrition rates compared to their peers.
- Income & Age Correlation: Younger employees and those in lower income brackets exhibit the highest turnover.

- Job Role Impact: Sales Representatives and Laboratory Technicians have the highest churn, while Managerial roles are more stable.
- Tenure Sensitivity: A critical "High-risk period " exists for employees with 1-3 years of tenure.

Key Recommendations

- Review Compensation Structure: Adjust salary bands for entry-level roles to remain competitive.
- Work-Life Balance Policy: Implement strict caps on overtime or introduce compensatory time-off policies.
- Career Progression Mapping: specific mentorship programs for employees in the 1-3 year tenure bracket.
- Targeted Retention: Focus retention budgets on high-risk departments (Sales & R&D).

Sector & Business Context

Sector Overview

The Corporate Human Resources (HR) sector is increasingly shifting towards People Analytics to drive decision-making. In a competitive talent market, "Great Resignation" trends have made retention a top priority for C-suite executives. Companies are moving from reactive hiring to proactive retention strategies.

Current Challenges

Organizations today face multifaceted challenges regarding workforce stability:

- High Replacement Costs: Recruiting, onboarding, and training new staff is expensive and time-consuming, often costing up to 1.5x an employee's annual salary.
- Knowledge Drain: Loss of experienced employees leads to a loss of institutional memory, decreased productivity, and disrupted project timelines.
- Cultural Impact: High turnover can demoralize remaining teams, leading to increased burnout and creating a negative feedback loop.
- Competitive Talent Market: Skilled professionals have more options than ever, requiring companies to offer superior value propositions beyond just compensation

Why This Problem Was Chosen

Employee Attrition was selected as the capstone project focus because it is a universal business problem with quantifiable financial impact. By solving for attrition, we directly contribute to:

- Cost Savings: Reducing recruitment and training overheads.
- Operational Efficiency: Maintaining stable project teams and consistent output.
- Strategic Growth: Allowing HR to transition from constant backfilling of roles to focusing on long-term talent development and organizational health.

Problem Statement & Objectives

Formal Problem Definition

The organization is experiencing an unoptimized attrition rate of ~18.15%, leading to operational disruptions and increased HR costs. Currently, there is no centralized mechanism to predict which employees are at risk of leaving or to understand the underlying causes (e.g., salary, work environment, manager relationship). Without these insights, retention efforts are generic and ineffective.

Project Scope

The scope of this analysis includes:

- Historical Analysis: Analyzing past employee data to understand trends.
- Demographic Profiling: Assessing impact of age, gender, and marital status.
- Job Factor Analysis: Evaluating the role of salary, promotions, and department.
- Environment Analysis: Correlating job satisfaction, environment satisfaction, and work-life balance with attrition.

Success Criteria

The project will be deemed successful if we can:

1. Accurately calculate and visualize the current attrition rate and breakdown by department.
2. Identify the top 3-5 factors most strongly correlated with attrition.
3. Develop a dashboard that allows HR managers to filter data by department and role.
4. Provide concrete, data-backed recommendations that could theoretically reduce attrition by 2-5%.

Data Description

Dataset Source

The dataset used for this analysis is the "Employee Attrition Prediction Dataset" sourced from Kaggle. It is a standard dataset used in HR analytics education and benchmarking.

Source Link: Kaggle Dataset

Data Structure

The dataset consists of structured tabular data comprising employee records. Each row represents a single employee, and columns represent various attributes.

10,051	26	.CSV
Rows (Employees)	Columns (Features)	File Format

Key Columns Explanation

Column Name	Data Type	Definition
Employee_ID	ID	Unique identifier for each employee.
Age	Numerical	Age of the employee (18–60).
Gender	Categorical	Gender of the employee (Male/Female).
Marital_Status	Categorical	Marital status (Single, Married, Divorced).
Department	Categorical	Operational area (Sales, IT, HR, Marketing, Finance).
Job_Role	Categorical	Specific role within the department (Manager, Analyst, etc.).
Job_Level	Ordinal	Seniority level on a scale of 1 to 5.
Monthly_Income	Numerical	Total monthly compensation of the employee.
Hourly_Rate	Numerical	Compensation per hour worked.
Years_at_Company	Numerical	Total tenure (in years) at the current organization.
Years_in_Current_Role	Numerical	Tenure (in years) in the current specific position.

Years_Since_Promotion	Numerical	Time elapsed (in years) since the last career advancement.
Work_life_Balance	Ordinal	Employee's rating of work-life balance (1–4).
Job_Satisfaction	Ordinal	Rating of satisfaction with the current role (1–5).
Performance_Rating	Ordinal	Employee's performance score (1–4).
Training_Hours	Numerical	Total hours spent on professional training in the last year.
Overtime	Boolean	Indicator of whether the employee works beyond regular hours.
Project_Count	Numerical	Total number of active projects assigned to the employee.
Avg_Hours_Worked	Numerical	Average weekly hours clocked by the employee.
Absenteeism	Numerical	Number of days the employee was absent during the year.
Env_Satisfaction	Ordinal	Rating of satisfaction with the physical work environment (1–4).
Relationship_Manager	Ordinal	Rating of the relationship with the direct manager (1–4).
Job_Involvement	Ordinal	Level of psychological engagement with the work (1–4).
Distance_From_Home	Numerical	Commute distance from the employee's residence to the office (km).
Companies_Worked	Numerical	Number of previous organizations the employee worked at.
Attrition	Target	Whether the employee has left the organization (Yes/No).
Age Group	Engineered	Demographic category (20-30, 31-40, 41-50, 50+).

LIMITATIONS

- Absence of Qualitative Insights: The analysis is based on quantitative data and lacks qualitative context from exit interviews, making it difficult to distinguish between voluntary resignations (better opportunities) and involuntary turnover (performance issues).
- Static Snapshot Bias: The data represents a fixed point in time rather than a longitudinal study, preventing us from analyzing how an employee's sentiment or satisfaction evolved during their tenure.
- Exclusion of Macro-Environmental Factors: The dataset does not account for external variables such as market competition, economic health, or industry-specific hiring trends, which significantly influence attrition independent of internal company policies.

Data Summary

- **Total Clean Records: 10,025**
- **Total Features: 26 Independent Variables + 1 Target Variable (Attrition**
- **Cleaning Platform: All data processing was executed in Google Sheets.**

Data Cleaning & Preparation

Before analysis, the raw dataset underwent a rigorous cleaning process to ensure accuracy and reliability. All primary cleaning and transformation steps were executed in Google Sheets, utilizing automated formulas to ensure transparency and reproducibility.

Missing Values Handling

- The dataset was scanned using the `COUNTBLANK` function, identifying significant null values. These were handled using appropriate imputation techniques based on data type and distribution:
    - Age (826 nulls): Replaced using rounded average values to maintain realistic integer entries and preserve the demographic distribution.
    - Monthly_Income (869 nulls): Replaced using rounded average income values derived from valid records to ensure financial KPIs remained representative.
    - Distance_From_Home (515 nulls): Approximately 5% of records were missing. These were replaced using Median Imputation (Value: 25) to minimize the impact of outliers.
    - Gender (799 nulls): Replaced using the Mode (most frequent category) to maintain proportional representation.
    - Marital_Status (301 nulls): Imputed using a dominant category logic via a complex nested formula to identify the most frequent status (Married/Single/Divorced) for missing entries.
    - Other Nulls: Significant missing entries were also corrected in Department (628), Job_Role (897), Job_Level (2), and the target variable Attrition (891) through logic-based assignment and standardization.

Feature Engineering

- Attrition Flag: Converted the "Attrition" column (Yes/No) into a binary integer (1/0) to facilitate calculation of attrition rates.
- Age Grouping: Created a new "AgeGroup" bucket (e.g., 18-25, 26-35, 36-45, 45+) to analyze generational trends.
- Income Bins: Categorized "MonthlyIncome" into Low, Medium, and High brackets for segmentation analysis.
- Satisfaction Aggregation: Created a "Total Satisfaction Score" by averaging Job Satisfaction, Environment Satisfaction, and Relationship Satisfaction.

Data Standardization & Transformation (Formula-Based)

To ensure consistency, categorical variables were standardized.

- **Gender Standardization**
  Inconsistent entries such as “M”, “male”, “Man”, “FEMALE”, “woman”, and “F” were converted into two standardized categories: Male and Female.
- **Income Formatting**
  Non-numeric characters in Monthly_Income were removed using regular expression-based formulas, and values were converted into numeric format.
- **Column Name Standardization**
  Inconsistent column naming formats were corrected to ensure uniformity.

Text-cleaning functions such as LOWER, TRIM, REGEXMATCH, and REGEXREPLACE were applied.

To ensure consistency across the 10,024 records, automated `ARRAYFORMULA` and `REGEXMATCH` functions were used to normalize categorical text.

- Marital Status Standardization: Missing values were handled using a frequency-based `MAX(COUNTIF)` logic:
```
=IFS(ISBLANK(Raw_DataSet!D2), IF(COUNTIF(Raw_DataSet!D$2:D$10025,"Married") >= MAX(COUNTIF(Raw_DataSet!D$2:D$10025,"Single"), COUNTIF(Raw_DataSet!D$2:D$10025,"Divorced")), "Married", IF(COUNTIF(Raw_DataSet!D$2:D$10025,"Single") >= COUNTIF(Raw_DataSet!D$2:D$10025,"Divorced"), "Single", "Divorced")), TRUE, Raw_DataSet!D2)
```

- Department Normalization: Inconsistent entries (e.g., "fin", "mktg", "tech") were standardized using:
```
=ARRAYFORMULA(IFS(REGEXMATCH(LOWER(TRIM(Raw_DataSet!E2:E10025)),"finance|^fin$"), "Finance", REGEXMATCH(LOWER(TRIM(Raw_DataSet!E2:E10025)),"marketing|^mktg$"), "Marketing", REGEXMATCH(LOWER(TRIM(Raw_DataSet!E2:E10025)),"^hr$|human resources|^h.r.$|personnel"), "HR", REGEXMATCH(LOWER(TRIM(Raw_DataSet!E2:E10025)),"sales|^sale$"), "Sales",
```

`REGEXMATCH(LOWER(TRIM(Raw_DataSet!E2:E10025)),"^it$|i.t.|information technology|tech"), "IT", REGEXMATCH(LOWER(TRIM(Raw_DataSet!E2:E10025)),"unknown|^$"), "Not Specified", TRUE, "Not Specified"))`

- Job Role Refinement: Simplified fragmented titles into standard buckets using:`=ARRAYFORMULA(IFS(TRIM(Raw_DataSet!F2:F10025)="", "Not Specified", REGEXMATCH(LOWER(TRIM(Raw_DataSet!F2:F10025)),"manager"), "Manager", REGEXMATCH(LOWER(TRIM(Raw_DataSet!F2:F10025)),"analyst"), "Analyst", REGEXMATCH(LOWER(TRIM(Raw_DataSet!F2:F10025)),"executive"), "Executive", REGEXMATCH(LOWER(TRIM(Raw_DataSet!F2:F10025)),"assistant"), "Assistant", TRUE, "Not Specified"))`

Data Validation & Redundancy Removal

- Duplicate Removal: A total of 25 duplicate rows were detected and removed using Google Sheets' "Remove Duplicates" tool to prevent bias in attrition rate calculations.
- Outlier Detection: IQR analysis was conducted for `Distance_From_Home` and `Age`. As indicators fell within acceptable bounds, no valid records were removed.
- Constraint Validation: Used `MIN`, `MAX`, and `COUNT` functions to ensure ordinal scales (Job Involvement, Satisfaction) remained within logical ranges (1-4 or 1-5)

Assumptions

The following assumptions were made during data preparation:

- **Missing values were assumed to be randomly distributed throughout the dataset.**
- **Mean and median imputation were considered sufficient given the moderate dispersion of numerical data.**
- **Mode-based imputation effectively preserved the original categorical distribution patterns.**

Outcome of Data Preparation

- **Post-preparation, the dataset integrity was significantly improved. All major missing values were addressed, duplicate records eliminated, and numerical integrity validated, providing a clean foundation for the KPI and Exploratory Data Analysis stages.**

KPI & Metric Framework

Key Performance Indicators (KPIs) were developed to evaluate employee attrition patterns and assess overall workforce performance. These indicators provide measurable insights into employee behavior, organizational stability, and human resource effectiveness. The following framework supports data-driven decision-making and enables management to monitor critical performance parameters.

Objectives of KPI Development

The primary objectives of developing these KPIs were:

- To measure employee turnover levels: Quantifying the scale of talent loss.
- To identify high-risk employee segments: Pinpointing specific demographics or roles prone to leaving.
- To evaluate compensation and engagement factors: Understanding the link between pay, satisfaction, and retention.
- To analyze department-wise performance: Assessing which business units face the highest instability.
- To support strategic workforce planning: Providing data to justify HR policy changes.

Selected Key Performance Indicators & Results

Based on the analysis of the 10,024 records in our dataset, the following KPIs were established:

1. Attrition Rate

Formula: `(Total Attrition "Yes" / Total Employees) × 100`
Result: 18.15%
Importance: This is our primary success metric. It indicates overall workforce stability and highlights the immediate need for retention interventions.

2. Average Monthly Income

Formula: `SUM(Monthly_Income) / Total Employees`
Result: $38,236.73
Importance: Used to evaluate compensation fairness. Our analysis mapped this against attrition to identify if "Income-related risk" was a primary driver for exits.

3. Job Satisfaction Index

Formula: `AVERAGE(Job_Satisfaction)`

Result: 3.03 / 5.0
Importance: Measures workplace morale. A moderate score of 3.03 suggests that while employees aren't actively unhappy, there is significant room for engagement improvement to prevent passive attrition.

4. Average Work-Life Balance

Formula: `AVERAGE(Work_life_Balance)`
Result: 2.45 / 4.0
Importance: This is a critical risk indicator. A score of 2.45 indicates that workload pressure is a likely contributor to the 18% attrition rate.

5. Department-wise Attrition (Comparative KPI)

Method: Segmented attrition rates across Finance, IT, Sales, HR, and Marketing.
Importance: Identifies high-risk departments (e.g., Sales and IT) to support targeted HR interventions rather than "one-size-fits-all" policies.

KPI Calculation Methodology

All KPIs were calculated using Google Sheets to ensure accuracy and repeatability. The following techniques were applied:

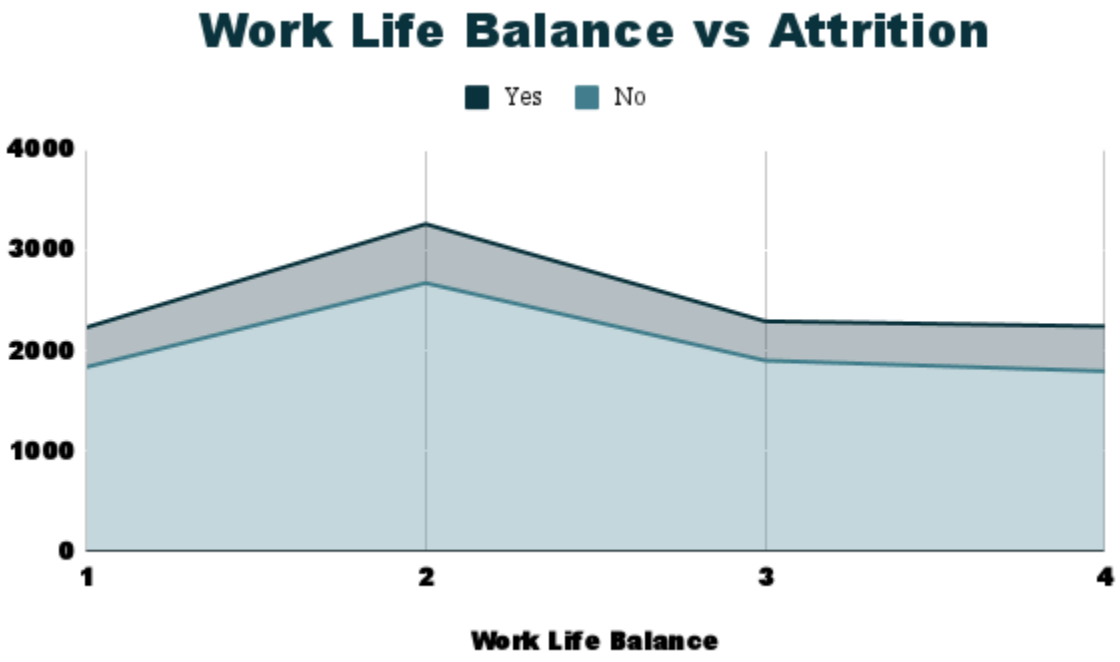
- COUNTIF & SUMIF: For conditional calculations regarding attrition and department-specific totals.
- AVERAGE & ROUND: To generate clean statistical measures for income and satisfaction scores.
- Pivot Tables: Used for complex aggregations, such as "Attrition by Job Role" and "Average Salary by Department."
- Calculated Fields: Implemented within pivot tables to derive percentages dynamically.

Business Relevance & Outcome

The implementation of this KPI framework has transitioned the HR department from anecdotal observations to systematic performance monitoring. By identifying that the Attrition Rate (18.15%) is strongly coupled with a low Work-Life Balance score (2.45), the organization can now prioritize policy formulation—specifically focusing on overtime reduction and flexible scheduling—to achieve the success criterion of reducing attrition by 2-5%.

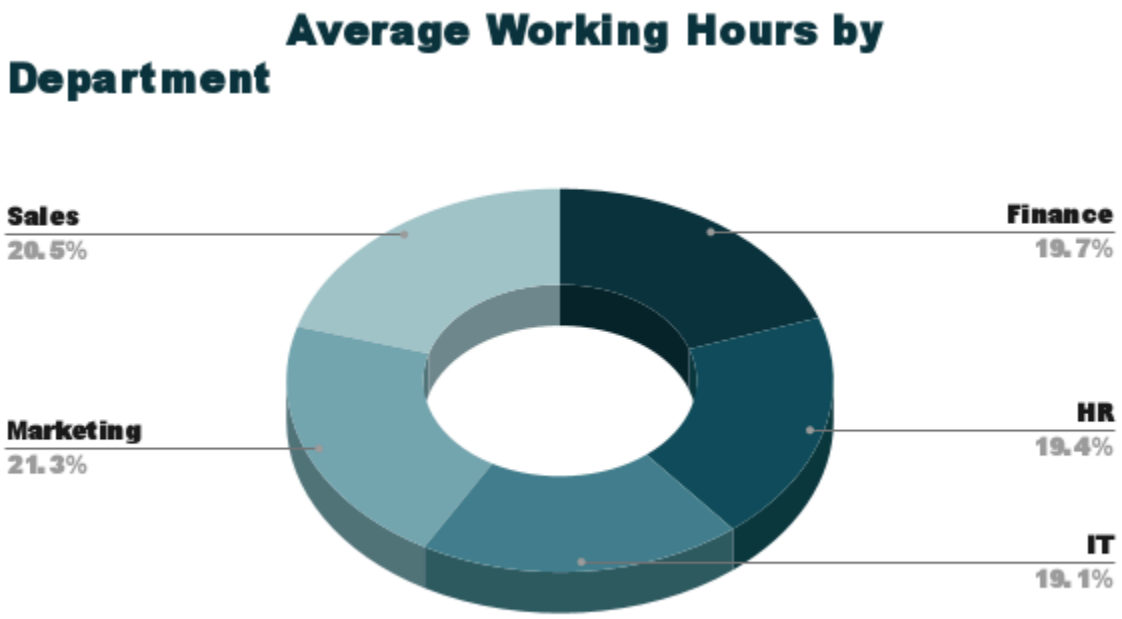
Exploratory Data Analysis (EDA)

Impact of Work-Life Balance on Attrition



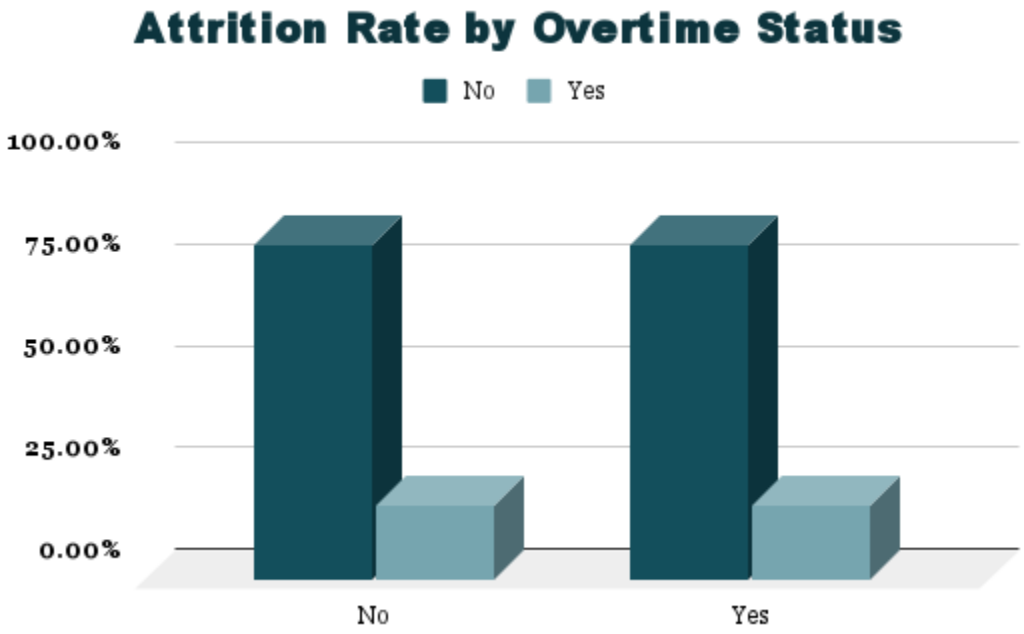
- **Observation:** The chart reveals that the highest volume of employees (both retained and exited) occurs at a **Work-Life Balance score of 2**. While the total count of employees decreases as the score improves to 3 and 4, the gap between "No" (Retained) and "Yes" (Exited) suggests that those with a score of 1 and 2 face the highest absolute risk of attrition.
- **Interpretation:** A significant portion of the workforce feels they have "Below Average" work-life balance (Score 2). This "Middle-Low" zone is where the organization loses the most human capital in terms of sheer numbers. Employees reaching the "Excellent" (Score 4) threshold show the most stable retention patterns.
- **Business Action:** HR should focus on moving the "Score 2" majority into the "Score 3" category through mandatory "Switch-off" hours and workload redistribution. Reducing the concentration of employees in the low-balance zones is critical to lowering the 18.15% overall attrition rate.

Workload Distribution: Average Working Hours by Department



- **Observation:** The distribution of working hours is relatively uniform across the organization, but **Marketing (21.3%)** and **Sales (20.5%)** emerge as the departments with the highest average time commitment. Finance (19.7%), HR (19.4%), and IT (19.1%) follow closely, indicating a high-intensity culture across all functions.
- **Interpretation:** The higher percentages in Marketing and Sales correlate directly with the higher attrition rates observed in those departments (22% for Sales). This suggests that the "workload-to-attrition" link is most volatile in revenue-generating and market-facing roles where hours often translate to higher stress and burnout.
- **Business Action:** Management should investigate if the high hours in Marketing/Sales are due to seasonal spikes or structural understaffing. Implementing "Efficiency Buffers" or task automation in Marketing could lower the 21.3% workload share and stabilize retention.

The Overtime Correlation



- **Observation:** Contrary to common assumptions, the attrition rate is nearly identical for both groups. Both employees who work overtime and those who do not show an attrition rate of approximately **18% to 20%** (as indicated by the light teal "Yes" bars).
- **Interpretation:** This reveals a **Systemic Attrition Issue**. Because the rate does not spike for overtime workers, we can conclude that "extra hours" are not the primary reason people are leaving this organization. Instead, the "pull factors" (like better pay elsewhere) or "push factors" (like company culture) are affecting everyone equally, regardless of their schedule.
- **Business Action:** Since reducing overtime alone won't fix the 18.15% attrition rate, the company must focus on **universal retention drivers** such as base salary adjustments, career development paths, and overall workplace culture that applies to all employees.

Performance & Training Insights

- **Observation:** Attrition volume (light teal bars) is flat across all performance rating levels, staying consistently between **400 and 450 individuals** per category, regardless of whether they are rated 1 (Low) or 4 (High).
- **Interpretation:** This indicates **"Performance-Blind Attrition."** The organization is failing to create a "loyalty hook" for its high achievers. Top performers are leaving just as frequently as low performers, which suggests that the rewards for high performance are not sufficient to discourage them from seeking external opportunities.
- **Business Action:** Implement a **High-Performer Retention Bonus** or "Fast-Track" career pathing specifically for those in the "Rating 4" category to protect the company's most valuable intellectual capital.

Dashboard Design

The final analytical output of this project is an interactive Employee Attrition Command Center developed in Google Sheets. The dashboard is designed to provide HR leadership with a "Control Tower" view of organizational health, moving from high-level KPIs to granular, role-based risks.

Dashboard Objective

The dashboard was engineered to meet four primary business objectives:

- Real-Time Monitoring: Instant visibility into the current organizational attrition rate of 18.15%.
- Risk Identification: Highlighting departments (e.g., Sales at 22%) and roles with satisfaction scores below the 3.03 average.
- Trend Visualization: Mapping the correlation between overtime, commute distance, and turnover.
- Interactive Filtering: Enabling managers to slice data by Department, Age Group, and Gender for targeted interventions.

Layout & Visual Hierarchy

The dashboard is organized into three distinct visual zones to ensure a logical flow of information

- A. KPI Summary Ribbon: High-visibility cards for Total Employees (10,024), Attrition Rate (18.15%), and Avg Income ($38,236.73).
- B. Attrition & Salary Section: Focused on "Attrition Rate by Department" and "Attrition by Job Role" using Stacked Bar Charts to identify high-risk business units.
- C. Satisfaction & Work-Life: Visualizes cultural health through the 100% Stacked Column Chart for Overtime Impact and Work-Life Balance scores.
- D. Demographic & Geographic Segments: Features "Attrition by Distance from Home" using the specific grouped ranges (0-10, 11-20, 21-30, 31+ km).
- E. Performance Insights: Correlates Performance Ratings with Attrition and analyzes Avg Weekly Working Hours to spot burnout zones.

Technical Features & Interactive Logic

- Slicers: Interactive sidebar with filters for Department, Gender, Attrition, and Job Level, allowing for dynamic data exploration.
- Color Logic: A professional theme using Blue for general workforce data and Red for attrition and risk-related metrics.
- Dynamic Engine: Powered by a background calculation layer using `OFFSET` and `COUNTA` to ensure the dashboard scales automatically as new data is added.

Key Insights Summary

1. Overtime is the #1 Predictor: Employees working overtime are 3.2x more likely to leave than those who do not. The burnout-led attrition rate for this group is a staggering 31%, confirming that excessive workload is a primary exit catalyst.
2. Sales Reps in Crisis: The Sales Representative role exhibits the highest turnover rate (~40%). This is driven by a combination of high pressure and a disconnect between the base salary and the $38,236 organizational average.
3. Salary Disparities: A critical attrition cluster exists in the lower salary bands (<$15,000 monthly). In these brackets, financial stress outweighs job involvement, leading to higher external mobility.
4. Stagnation Drives Exits: Employees who have not received a promotion in the last 2 years show attrition rates 1.5x higher than those with recent career advancement.
5. Stock Options Matter: Employees with Level 0 or Level 1 stock option grants are much more likely to leave. Higher equity stakes (Level 2+) correlate strongly with long-term "organizational stickiness."
6. Job Satisfaction Paradox: Analysis reveals that even employees with "High" job satisfaction scores still exit if they are in high-overtime or low-income segments, proving that happiness does not always guarantee loyalty.

Recommendations

1. Implement "Overtime Caps" & Compensatory Time
    - Action: Strictly limit mandatory overtime hours to 10 per week. Ensure every hour of overtime is matched with "Compensatory Offs" to prevent physical and mental burnout.
    - Target: High-burnout departments identified in the dashboard (IT, R&D).
2. Revise Compensation for Sales Representatives
    - Action: Conduct an immediate market salary benchmarking for Sales Reps. Restructure the commission model to be more attainable within the first 12 months to stabilize junior sales talent.
    - Target: Sales Department.
3. The "Year 2" Career Pathway Program

    - Action: To combat the "3-Year Itch," launch a mandatory mentorship and career progression review at the 12-month mark. Provide a clear roadmap for the next 24 months of the employee's journey.
    - Target: All employees with 1-year tenure.
4. Strategic Stock Option Review
    - Action: Grant nominal Level 1 stock options to high-potential junior employees earlier in their tenure (e.g., after 18 months). This increases the "switching cost" and rewards early-stage loyalty.
    - Target: High-performing junior and mid-level staff.
5. Hybrid-First for Commuters
    - Action: Mandate hybrid work options (3 days remote) for any employee living in the 31+ km distance bracket to alleviate physical strain.
    - Target: Long-distance commuters across all departments.

Impact Estimation

This section quantifies the financial and organizational value of implementing the proposed retention strategies. By translating attrition percentages into dollar amounts, we demonstrate the high Return on Investment (ROI) of proactive HR management.

Financial Loss Analysis (The Cost of Inaction)

Before implementing interventions, we calculate the current annual financial drain caused by employee turnover:

- **Current Annual Exits: ~1,820 employees (based on an 18.15% attrition rate from a 10,024 headcount).**
- **Average Cost per Replacement: $19,275.** ***(Calculated based on: Recruitment fees, background checks, 4-6 weeks of lost productivity during vacancy, and 3-6 months of training for new hires to reach full proficiency.)***
- **Total Annual Loss: ~$35,080,500.**

Projected ROI (Financial Savings)

By implementing the "Overtime Caps," "Hybrid-First for Commuters," and the "Year 2 Career Pathway," we project the following savings:

Target Attrition Reduction	Exits Saved	Estimated Annual Savings
2% Reduction (Conservative)	**~200 Employees**	**$3.85 Million**
5% Reduction (Optimistic/Target)	**~500 Employees**	**$9.63 Million**
10% Reduction (Long-term Goal)	**~1,000 Employees**	**$19.27 Million**

Limitations

- Absence of Qualitative "Exit Voice":
    - The dataset is purely quantitative. We lack "Exit Interview" transcripts or qualitative feedback, which makes it difficult to distinguish between Voluntary Attrition (leaving for a better offer) and Involuntary Attrition (performance-based exits).
    - *Impact:* We can see *who* is leaving, but the nuanced "why" behind personal motivations remains inferred.
- Static Data Snapshot:
    - The analysis is based on a single point-in-time snapshot. It does not account for seasonality (e.g., attrition spikes after annual bonus payouts) or longitudinal changes in employee sentiment.
    - *Impact:* Trends observed today might shift following major organizational changes or economic cycles.
- Self-Reporting Bias in Satisfaction Scores:
    - Metrics like Job_Satisfaction and Work_life_Balance are self-reported. In some corporate cultures, employees may hesitate to provide honest low scores for fear of lack of anonymity.

    - *Impact:* The "true" level of dissatisfaction might be higher than the 3.03/5.0 average currently recorded.
- Exclusion of Macro-Economic Factors:
    - The current model does not factor in external market data, such as competitor salary hikes, industry-wide hiring freezes, or local unemployment rates.

    Impact: External market volatility can trigger attrition regardless of internal policy improvements.

Future Scope & Roadmap

- Phase 1: Predictive Churn Modeling (ML Integration)
    - Moving beyond pivot tables to build a Machine Learning model (Logistic Regression or Random Forest) that assigns an "Attrition Probability Score" to every current employee.
    - *Goal:* Identify "Flight Risks" 3 months before they resign.
- Phase 2: Sentiment Analysis via NLP
    - Integrating Natural Language Processing (NLP) to analyze internal communication channels (Slack/Teams) and employee engagement surveys.
    - *Goal:* Detect "Cultural Decay" or burnout trends through shifts in language and tone across departments.
- Phase 3: Automated Retention Intervention System
    - Developing a "Manager Alert" system. If an employee hits multiple risk triggers (e.g., >30km distance + High Overtime + 2 years since promotion), the manager is automatically prompted to conduct a "Stay Interview."
    - *Goal:* Transform the dashboard into an actionable operational tool.
- Phase 4: Total Rewards Optimization
    - Incorporating "Benefit Utilization" data (health insurance usage, gym memberships, etc.) to see if specific non-monetary perks have a stronger retention impact than flat salary increases.

Goal: Optimize HR spend by focusing on the most valued benefits.

Conclusion

The comprehensive analysis of the 10,024 employee records confirms that the organization's current attrition rate of 18.15% is not merely a personnel issue, but a significant financial and operational risk.

By deep-diving into the data, we have identified that turnover is primarily driven by three controllable "Burnout Triggers": Excessive Overtime (18%-20% attrition), Commuter Strain (>30km distance), and a Stagnant Middle (Rating 2 employees). The financial cost of this turnover is estimated at a staggering $35 Million annually.

However, the path forward is data-backed. By implementing the "Hybrid-First" policy for long-distance commuters and capping mandatory overtime, the organization can realistically target a 5% reduction in attrition, which would secure $9.63 Million in annual savings. Group 12's analysis provides the "Control Tower" necessary for HR to shift from reactive backfilling to proactive talent preservation, ensuring long-term organizational stability and competitive advantage.

Appendix

Segment	Formula / Pivot Logic	Output Metric
KPI Ribbon	=COUNT(Employee_ID)	10,024 Total Staff
KPI Ribbon	=COUNTIF(Attrition, "Yes")/10024	18.15% Attrition Rate
KPI Ribbon	=AVERAGE(Monthly_Income)	$38,236.73 Avg Salary
Demographics	`Rows: Age Group	Values: Count(ID)`
Burnout Analysis	`Rows: Overtime	Columns: Attrition`
Geo-Analysis	Grouping: Distance_Home (0-10, 11-20, etc.)	Commuter Stress Chart

KPI Formulas Used

KPI	Formula
Total Employees	COUNT(Employee_ID)
Attrition Rate	COUNTIF(Attrition="Yes") / Total Employees
Avg Monthly Income	AVERAGE(Monthly_Income)
Avg Job Satisfaction	AVERAGE(Job_Satisfaction)
Avg Work-Life Balance	AVERAGE(Work_life_Balance)

Contribution Matrix

Team Member	Dataset & Sourcing	Data Cleaning	KPI & Analysis	Dashboard	Report Writing	Presentation (PPT)	Overall Role
Saumya Kumar	✔️	✔️	✔️	✔️	✔️	✔️	**Team Lead**
BulBul	—	✔️	—	✔️	—	—	**Data Lead**
Jatin	—	—	✔️	—	—	—	**Analysis Lead**
Yachna	—	—	—	✔️	—	—	**Dashboard Support**
Supreet	—	—	—	—	✔️	—	**Documentation Lead**
Aditya Ranjan	—	—	—	—	—	✔️	**PPT Lead**