

HR Analytics Dashboard Pipeline

TABLE OF CONTENTS:

- [Project Overview](#project-overview)
- [Architecture](#architecture)
- [Tech Stack](#tech-stack)
- [Dataset](#dataset)
- [Project Structure](#project-structure)
- [dbt Models](#dbt-models)
- [Key Metrics & KPIs](#key-metrics--kpis)
- [Power BI Dashboards](#power-bi-dashboards)
- [Setup Instructions](#setup-instructions)
- [Maintenance & Versioning](#maintenance--versioning)

PROJECT OVERVIEW:

This project builds a comprehensive HR analytics pipeline that ingests raw employee data, transforms it through multiple layers, calculates key HR metrics and risk scores, and visualizes insights through interactive dashboards. The purpose was to analyze employee attrition, engagement, compensation, and risk factors to support HR decision-making.

ARCHITECTURE:

Raw Data (CSV)



PostgreSQL Database (raw_data schema)



dbt Transformations (3 layers)

- └── Staging: Data cleaning & standardization
- └── Intermediate: KPI calculations & risk scoring
- └── Marts: Business-ready fact & dimension tables



Power BI Dashboards (Interactive visualizations)

TECH STACK:

- **Database:** PostgreSQL 17
- **Data Transformation:** dbt (Data Build Tool) 1.10.13
- **Visualization:** Microsoft Power BI

- **Version Control:** Git/GitHub
- **Languages:** SQL, DAX

DATASET:

- **Source:** [IBM HR Analytics Employee Attrition Dataset](#)
- **Records:** 1,520 employees (1,470 original + 50 added for maintenance demonstration)
- **Columns:** 28 attributes (demographics, compensation, engagement, performance)
- **Size:** ~500KB

DATA DICTIONARY:

Demographics:

- employee_number: Unique employee identifier
- age: Employee age in years
- gender: Male/Female
- education_level: Education background (1-5 scale)
- marital_status: Single/Married/Divorced

Employment:

- department: Sales, Research & Development, Human Resources
- job_role: Specific position title
- years_at_company: Tenure with organization
- years_in_current_role: Time in current position
- years_since_last_promotion: Promotion recency

Compensation & Performance:

- monthly_income: Employee salary (monthly)
- annual_income: Calculated yearly salary
- performance_rating: 3 (Outstanding) or 4 (Excellent)
- percent_salary_hike: Recent raise percentage

Engagement & Satisfaction (1-4 scale):

- job_satisfaction_level: Low/Medium/High/Very High
- environment_satisfaction_level: Workplace satisfaction
- work_life_balance_level: Bad/Good/Better/Best
- job_involvement_level: Engagement with work

Training & Development:

- training_times_last_year: Number of training sessions (0-6+)

- over_time: Yes/No - Works overtime

Attrition:

- attrition: Yes/No - Has employee left
- attrition_flag: 1 (left) or 0 (current employee)

PROJECT STRUCTURE:

```

hr_analytics/
└── models/
    ├── sources.yml           ##Data source definitions
    ├── staging/
    │   └── stg_hr_employees.sql      ##Raw data cleaning & standardization
    ├── intermediate/
    │   └── int_employee_metrics.sql    ##KPI calculations
    │   └── int_employee_risk_scoring.sql ##Attrition risk model
    └── marts/
        ├── fct_employee_metrics.sql    ##Fact table (all metrics)
        └── dim_employees.sql          ##Dimension table (employee attributes)
    └── dbt_project.yml           ##dbt configuration
    └── .dbt/profiles.yml         ##Database connection config
    └── README.md                 ##This file
    └── MAINTENANCE.md           ##Version history & updates
    └── dashboards/
        └── HR_Analytics_Dashboard.pbix ## Power BI dashboard file

```

DBT MODELS:

Staging Layer

- stg_hr_employees:
 - Cleans and standardizes raw data
 - Removes null records
 - Standardizes text fields (TRIM, UPPER)
 - Converts coded values to human-readable labels
 - Calculates derived fields (annual_income)

Intermediate Layer

- int_employee_metrics:
 - Calculates business KPIs
 - Age groups (categorization)

- Tenure categories (New, Established, Experienced, Tenured) with sort order
 - Income brackets (\$40K-\$60K, etc.)
 - Engagement Score (average of satisfaction metrics: 1-4 scale)
 - Promotion status (Promoted This Year, Due for Review, etc.)
 - Training investment levels and scores (0-3 scale)
- int_employee_risk_scoring
 - Predictive attrition risk model
 - Calculates risk factors (0-2 points each):
 - Satisfaction risk
 - Work-life balance risk
 - Promotion recency risk
 - Overtime work
 - Compensation risk
 - Performance risk
 - Risk Score:
 - Sum of all factors (0-11 scale)
 - Risk Categories with sort order:
 - Critical Risk (6+ points)
 - High Risk (4-5 points)
 - Medium Risk (2-3 points)
 - Low Risk (0-1 points)

Marts Layer

- fct_employee_metrics
 - Fact table with all employee metrics
 - Combines employee attributes with calculated KPIs
 - Joins metrics with risk scoring
 - Includes training investment score
 - Ready for Power BI analysis
- dim_employees
 - Dimension table for filtering
 - Employee attributes (department, job role, education, gender, etc.)
 - Age grouping
 - Tenure categorization

KEY METRICS & KPI:

Attrition Analysis:

- Attrition Rate %: Calculated as (Count of employees who left / Total employees) × 100
 - DAX Formula: `DIVIDE(CALCULATE(COUNT([employee_number]), [attrition_flag] = 1), COUNT([employee_number])) * 100`
- Attrition by Department: Identify high-turnover areas
- Risk Category Distribution: Count of employees in each risk tier

Engagement Metrics:

- Engagement Score (1-4): Average of job satisfaction, environment satisfaction, work-life balance, job involvement
- Department Engagement Comparison
- Satisfaction Trends by Demographic

Compensation & Tenure:

- Average Income by Department
- Income Distribution by Tenure
- Promotion Recency: Years since last promotion
- Tenure Categories: Employee distribution

Performance & Development

- Performance Rating Distribution
- Training Investment Score (0-3 scale)
- Training Level Categories
- Correlation between training and attrition/performance

DATA QUALITY & TESTING:

dbt Tests Implemented:

- Source freshness: Validates data is up-to-date
- Not null tests: Ensures critical fields (employee_number, department) have values
- Unique tests: Validates employee_number is unique
- Referential integrity: Checks joins between models

Data Validation:

- All 1,520 employee records processed successfully
- Zero duplicate employee numbers
- All risk scores calculated correctly (0-11 range)
- Engagement scores validated (1-4 scale)

POWER BI DASHBOARDS OVERVIEW:

The dashboard consists of 5 interactive pages:

1. Executive Overview

- Total employee count and attrition rate KPIs
- Department breakdown
- High-level metrics for quick insights

2. Employee Retention & Risk Analysis

- Risk category distribution
- Attrition analysis by department and demographics
- Risk factor breakdown (satisfaction, work-life balance, compensation, etc.)
- Interactive filtering by risk level

3. Compensation & Tenure

- Income distribution across departments and roles
- Tenure analysis
- Promotion status tracking
- Compensation vs. attrition correlation

4. Department Deep Dive

- Department metrics (employee count, average annual income, average engagement score, etc.)
- Performance distribution by department
- Filter by department, education level, gender

5. Training Analysis (Enhancement - v1.2)

- Training investment score by department
- Training distribution matrix
- Correlation between training and performance
- Training vs. attrition analysis

Features:

- Interactive slicers for filtering by department, education, gender, promotion status
- Drill-through capabilities for detailed analysis
- KPI cards for key metrics
- Consistent color schemes and formatting

SETUP INSTRUCTIONS:

Prerequisites:

- PostgreSQL 17+

- Python 3.10+
- dbt-postgres adapter
- Power BI Desktop
- Git

Installation:

1. Clone the repository

```
git clone https://github.com/yourusername/hr-analytics-dbt.git
cd hr-analytics-dbt
```

2. Set up dbt

```
pip install dbt-core dbt-postgres
dbt debug
```

3. Load raw data

- Import CSV to PostgreSQL (using PGAdmin or psql)
- Create `raw_data.hr_employees` table and load CSV

4. Run dbt transformations

```
dbt run
dbt test
```

5. Open Power BI dashboard

- Connect to PostgreSQL database (localhost, `hr_analytics`)
- Load `staging.fct_employee_metrics` and `staging.dim_employees`
- Open `dashboards/HR_Analytics_Dashboard.pbix`

DBT COMMANDS:

- Run all models: **dbt run**
- Run specific model: **dbt run --select stg_hr_employees**
- Test data quality: **dbt test**
- Generate documentation: **dbt docs generate**
- Check connection: **dbt debug**

LIMITATIONS & FUTURE IMPROVEMENTS:

Current Limitations:

- Snapshot data: Single point-in-time view; no historical trend analysis
- Synthetic data: Based on IBM sample dataset, not real company data

- Limited attributes: Missing some key HR metrics (hire date, manager relationship, department transfers)
- No predictive modeling: Risk scoring is rule-based, not ML-powered

Future Enhancements:

- Add time-series analysis with monthly snapshots
- Implement machine learning for attrition prediction
- Include salary benchmarking against industry standards
- Add manager effectiveness metrics
- Create employee journey visualization

MAINTENANCE & VERSIONING:

This project demonstrates data pipeline maintenance and enhancement capabilities. See MAINTENANCE.md for detailed version history.

Current Version: 1.2

Recent Updates:

- v1.1: Added 50 new employee records to demonstrate data refresh process
- V1.2: Enhanced with training investment scoring and analysis page

Author: Sierra Zuzick

Last Updated: 11/03/2025