



Relatório Final do Projeto

Análise de Dados de RH – IBM



- Apresentação do Projeto
 - Dicionário de Dados
 - Análise Exploratória
 - Normalização e Modelo Relacional
 - Scripts T-SQL
 - Caso de Uso NoSQL
 - Transformação de Dados
 - Modelo OLAP
 - Métricas e Visualizações
 - Conclusões e Recomendações
-

Índice

Apresentação do Projeto

Objetivo

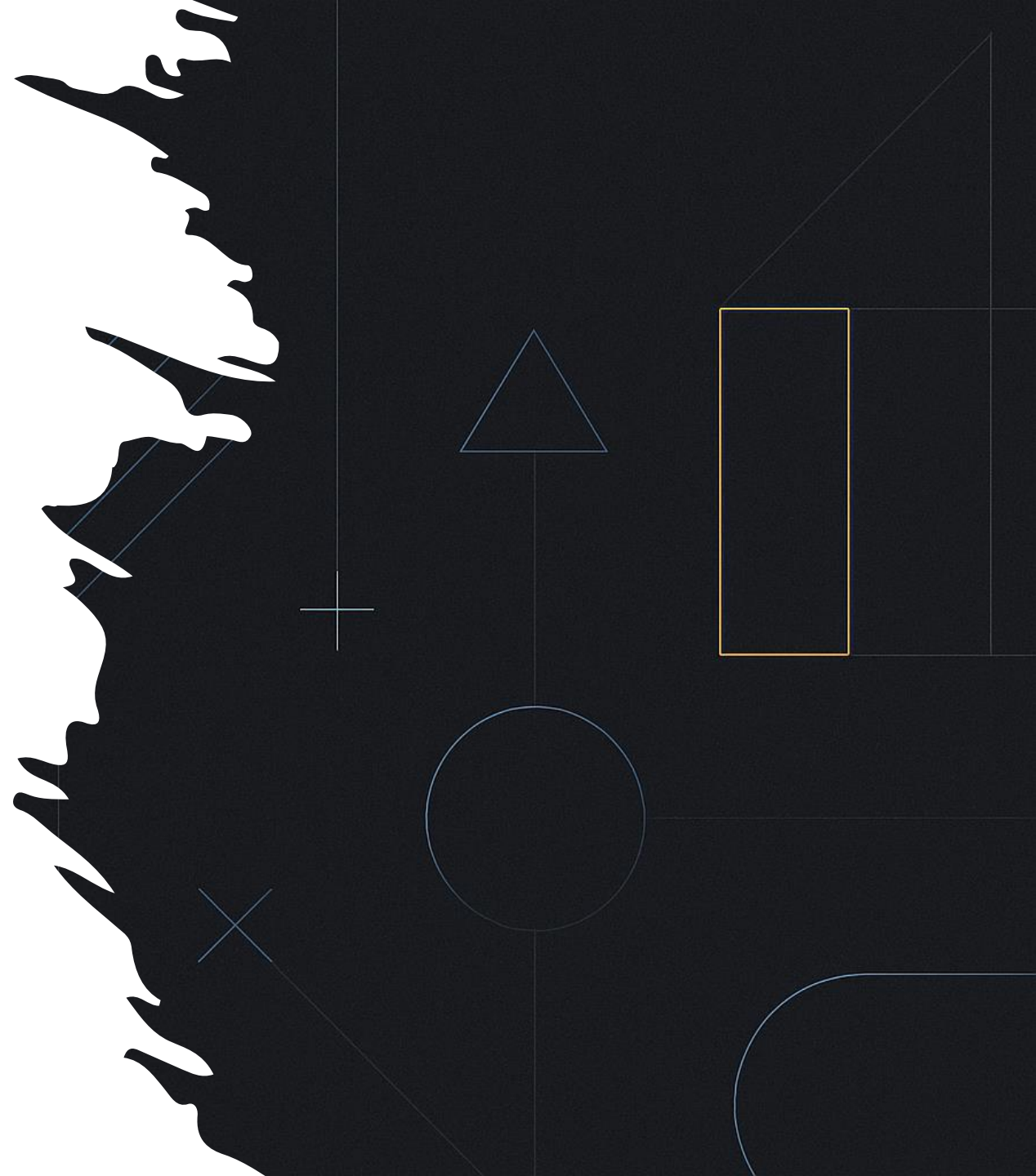
Análise de dados de RH para identificar fatores de rotatividade

Dataset:

- Dados de funcionários da IBM com informações demográficas, salariais e de desempenho

Metodologia:

- SQL Server para modelagem relacional
- MongoDB para caso NoSQL
- Power BI para transformação, visualização e análise



Key	Column	Data Type	Allow Nulls	Properties	Descrição
1	PK Job_History_ID	INT		Identity (1,1)	Número de identificação único para o histórico de trabalho.
	TotalWorkingYears	TINYINT			Anos totais de trabalho dos funcionários.
	YearsInCurrentRole	TINYINT			Anos de trabalho dos funcionários na função atual.
	YearsSinceLastPromotion	TINYINT			Os anos desde a última promoção dos funcionários.
	YearsWithCurrManager	TINYINT			Os anos que os funcionários trabalharam com o gestor atual.
	YearsAtCompany	TINYINT			Os anos que o funcionário trabalhou para a IBM.
	NumCompaniesWorked	TINYINT			O número de empresas para as quais o funcionário trabalhou antes.
	FK Employee_ID				(Chave estrangeira para a entidade Employee)
10	PK DepartmentID	INT		Identity (1,1)	Número de identificação único para o departamento da IBM.
	Department	nvarchar(50)			O nome do departamento da IBM. (Ex: Human Resources / Research & Development / Sales)
10	PK PerformanceID	INT		Identity (1,1)	Número de identificação único para o desempenho dos funcionários.
	PerformanceRating	TINYINT			Uma classificação de 0 a 4 que indica o quão bem um funcionário trabalha. 1 sendo baixo e 4 sendo muito alto.
	WorkLifeBalance	TINYINT			Uma classificação de 0 a 4 que indica o quão bem um funcionário equilibra o seu trabalho e vida. 1 sendo baixo e 4 sendo muito alto.
10	FK Employee_ID				(Chave estrangeira para a entidade Employee)
2	PK JobRoleID	INT		Identity (1,1)	Número de identificação para uma determinada função.
	JobRole	varchar(50)			O título da função.
3	PK OvertimeID	INT		Identity (1,1)	(Número de identificação para a indicação de horas extras)
	OverTime	VARCHAR(3)			Indica se o funcionário trabalha horas extras ou não. (Yes / No)
11	PK JobID	INT		Identity (1,1)	(Número de identificação para os detalhes do trabalho)
	JobInvolvement	TINYINT			Uma medida de quão envolvido um funcionário está no seu trabalho. 1 sendo baixo e 4 sendo muito alto.
	JobLevel	TINYINT			Uma medida de quão valioso é o trabalho. (1-5, onde 1: Funcionário, 5: Executivo)
	FK JobRole				(Chave estrangeira para a entidade JobRole)
	FK OverTime				(Chave estrangeira para a entidade Overtime)
4	PK BuisenessTravelID	INT		Identity (1,1)	(Número de identificação para os detalhes de viagem de negócios)
	BusinessTravel	VARCHAR(50)			Com que frequência um funcionário viaja a negócios. (Travel Frequently / Travel Rarely / Non-Travel)
12	PK DistanceID	INT		Identity (1,1)	Número de identificação para determinadas distâncias.
	DistanceFromHome	INT			Distância de casa dos funcionários em milhas.
	FK Employee_ID				(Chave estrangeira para a entidade Employee)
12	FK BuisenessTravelID				(Chave estrangeira para a entidade Buiseness)
5	PK GenderID	INT		Identity (1,1)	(Número de identificação para o gênero)
	Gender	VARCHAR(10)			Fornece o gênero dos funcionários. (Feminino/Masculino).
6	PK MaritalStatusID	INT		Identity (1,1)	(Número de identificação para o estado civil)
	MaritalStatus	VARCHAR(10)			Indica o estado civil dos funcionários. (Divorced / Married / Single)
13	PK Personal_ID	INT		Identity (1,1)	Número de identificação para certas informações pessoais.
	Age	INT			Fornece a idade dos funcionários.
	FK Gender				(Chave estrangeira para a entidade Gender)
	FK MaritalStatus				(Chave estrangeira para a entidade MaritalStatus)
	FK Employee_ID				(Chave estrangeira para a entidade Employee)
14	PK Salary_ID	INT		Identity (1,1)	Número de identificação para as informações de salário de cada funcionário.
	HourlyRate	INT			Salário por hora do funcionário.
	MonthlyRate	INT			Salário por mês do funcionário.
	MonthlyIncome	INT			Rendimento mensal do funcionário.
	PercentSalaryHike	TINYINT			Porcentagem de aumento salarial do funcionário.
	StockOptionLevel	TINYINT			Capacidade de comprar ações da empresa por um período específico. (0-3)
	DailyRate	TINYINT			Taxa de salário diário.
	FK Employee_ID	INT			(Chave estrangeira para a entidade Employee)
15	PK Satisfaction_ID	INT		Identity (1,1)	Número de identificação para as informações de satisfação de cada funcionário.
	EnvironmentSatisfaction	TINYINT			Fornece informações sobre o quão satisfeitos os funcionários estão com o seu ambiente. 1 sendo baixo e 4 sendo muito alto.
	RelationshipSatisfaction	TINYINT			Fornece informações sobre o quão satisfeitos os funcionários estão com o seu relacionamento. 1 sendo baixo e 4 sendo muito alto.
	JobSatisfaction	TINYINT			Fornece informações sobre o quão satisfeitos os funcionários estão com o seu trabalho. 1 sendo baixo e 4 sendo muito alto.
	FK Employee_ID				(Chave estrangeira para a entidade Employee)
7	PK AttritionID	INT		Identity (1,1)	(Número de identificação para o status de atrito)
	Attrition	VARCHAR(10)			Menciona se o funcionário se demitiu ou não. (Current Employee / Voluntary Resignation / Termination)
16	PK Employee_ID	INT		Identity (1,1)	Número de identificação para cada funcionário.
	FK Attrition				(Chave estrangeira para a entidade Attrition)
	FK JobID				(Chave estrangeira para a entidade JOB)
	FK DepartmentID				(Chave estrangeira para a entidade Department)
8	PK EducationField	INT		Identity (1,1)	(Número de identificação para a área de formação)
	EducationField	VARCHAR(50)			Fornece informações sobre a área de formação do funcionário.
17	PK Education_ID	INT		Identity (1,1)	Número de identificação para as informações de educação de cada funcionário.
	Education	TINYINT			Fornece informações sobre o nível de escolaridade do funcionário. (1: Abaixo do ensino superior, 5: Doutramento)
	TrainingTimesLastYear	TINYINT			Número de vezes que o funcionário foi treinado no último ano.
	FK Employee_ID				(Chave estrangeira para a entidade Employee)
	FK EducationField				(Chave estrangeira para a entidade EducationField)

Dicionário de Dados

Principais entidades:

- HISTORY
- DEPARTMENT
- PERFORMANCE
- JOB
- DISTANCE
- PERSONAL INFORMATION
- SATISFACTION
- EMPLOYEE DETAILS
- EDUCATION

Resumo da Análise Exploratória dos Dados

Visão Geral do Dataset

O dataset analisado contém dados de funcionários da IBM, com informações detalhadas sobre carreira, desempenho, satisfação, dados demográficos e histórico profissional. A análise exploratória revelou insights importantes sobre a composição da força de trabalho e fatores relacionados à rotatividade.

Principais Correlações Identificadas

- Forte correlação negativa entre satisfação no trabalho e attrition
- Correlação moderada entre horas extras e attrition
- Correlação positiva entre tempo na empresa e nível salarial
- Correlação fraca entre idade e satisfação no trabalho
- Correlação significativa entre distância de casa e attrition

Conclusões Preliminares

- O departamento de Sales apresenta a maior taxa de rotatividade
- Funcionários que fazem horas extras têm maior probabilidade de deixar a empresa
- A baixa satisfação no trabalho é um forte indicador de potencial saída
- Funcionários com menos de 2 anos de empresa representam uma parcela significativa das saídas
- O perfil etário predominante (30-40 anos) apresenta taxa de attrition inferior à média geral

Processo de normalização, com a apresentação da 3NF

Para garantir uma base de dados eficiente, coesa e preparada para análise, foi realizada a normalização do dataset IBM HR Attrition. A normalização visa reduzir a redundância de dados, melhorar a integridade e facilitar consultas mais flexíveis, especialmente quando os dados serão usados para análises exploratórias e visualizações.

Objetivos principais da normalização aplicada



Objetivos principais da normalização aplicada

1- Eliminar redundância de dados:

Informações como *Departement*, *Marital Status*, *Job Role* e *Education* aparecem repetidas varias vezes na tabela original. Essas informações foram movidas para tabelas separadas.

2- Melhorar a integridade referencial:

Ao separar os dados em tabelas distintas com relações claras via chaves estrangeiras, garantimos que apenas valores válidos e consistentes sejam utilizados.

3- Facilitar consultas e agregações:

Com a estrutura normalizada, torna-se mais simples aplicar joins para analyses como por exemplo a rotatividade por cargo ou departamento.

4- Preparar a base para o uso em PowerBI:

Separar atributos categóricos em dimensões permite criar dashboards mais performáticos.



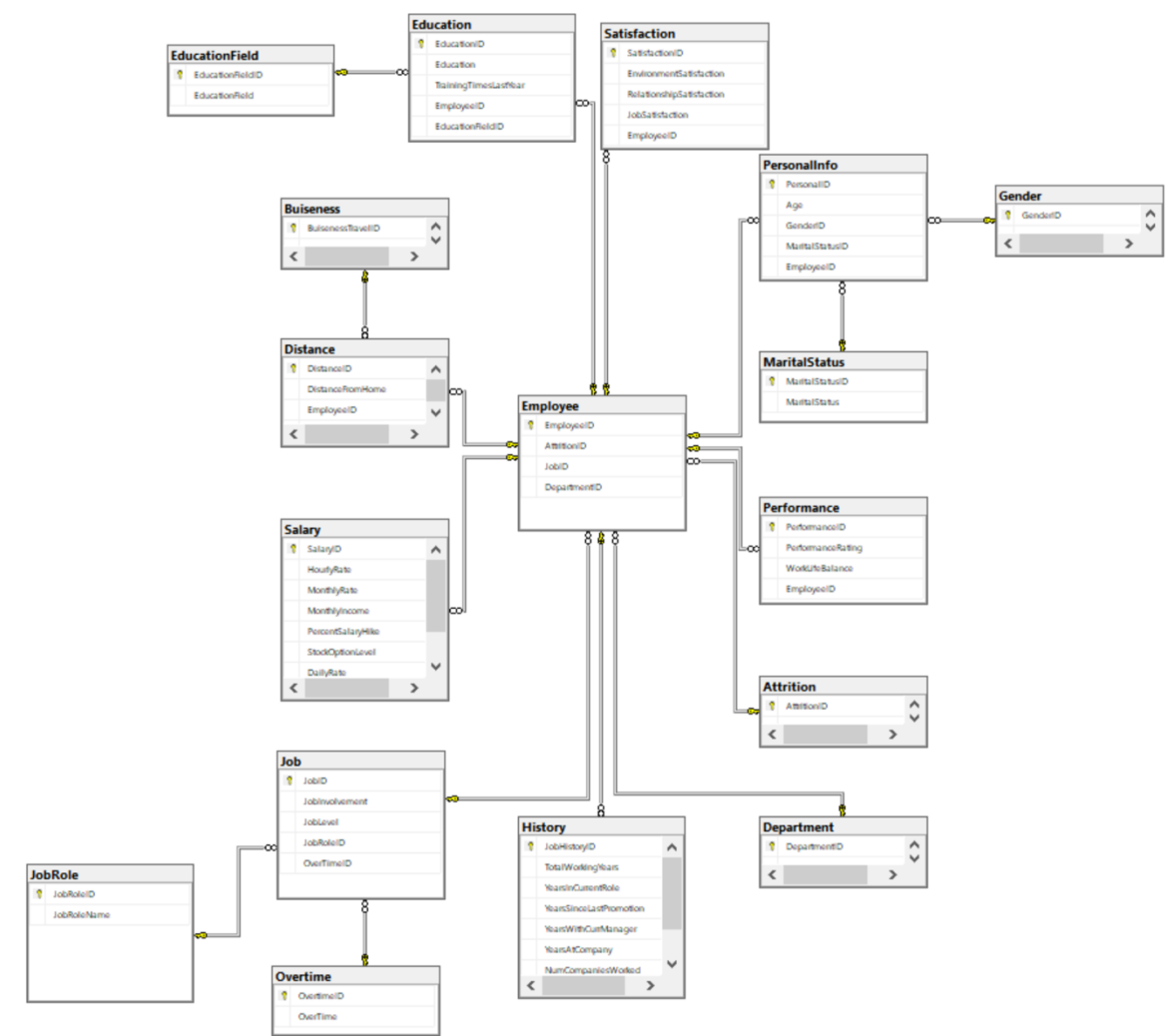


Imagem do diagrama ER do modelo relacional proposto

Excertos dos principais scripts T-SQL

```
--relatório de rotatividade de funcionários (attrition) por departamento,
--mostrando quantos saíram (Yes) e quantos permaneceram (No)
SELECT
    d.department AS Department,
    SUM(CASE WHEN a.attrition = 'Yes' THEN 1 ELSE 0 END) AS Yes,
    SUM(CASE WHEN a.attrition = 'No' THEN 1 ELSE 0 END) AS No
FROM
    employee e
JOIN
    department d ON e.departmentid = d.departmentid
JOIN
    attrition a ON e.attritionid = a.attritionid
GROUP BY
    d.department
ORDER BY
    d.department;

-- Relatório de estatísticas salariais por departamento,
--considerando apenas os funcionários que ainda estão na empresa (attrition = 'No').

SELECT
    d.department AS Departamento,
    FORMAT(ROUND(AVG(s.monthlyincome), 2), 'N2', 'pt-PT') + ' €' AS MédiaSalarial,
    FORMAT(MAX(s.monthlyincome), 'N2', 'pt-PT') + ' €' AS SalárioMáximo,
    FORMAT(MIN(s.monthlyincome), 'N2', 'pt-PT') + ' €' AS SalárioMínimo
FROM
    employee e
JOIN
    department d ON e.departmentid = d.departmentid
JOIN
    attrition a ON e.attritionid = a.attritionid
JOIN
    salary s ON e.employeeid = s.employeeid
WHERE
    a.attrition = 'No'
GROUP BY
    d.department
ORDER BY
    d.department,
    MédiaSalarial
DESC;
```

Excertos dos principais scripts T-SQL

-- relatório por departamento mostrando a quantidade de funcionários que fazem ou não horas extras.

```
SELECT
    d.department AS Departamento,
    o.overtime AS OverTime,
    COUNT(*) AS Totalfuncionarios
FROM
    employee e
JOIN
    department d ON e.departmentid = d.departmentid
JOIN
    job j ON e.jobid = j.jobid
JOIN
    overtime o ON j.OverTimeID = o.OvertimeID
GROUP BY
    d.department,
    o.overtime
ORDER BY
    d.department,
    o.overtime;
```

--relatório que mostra a média de idade dos funcionários em cada departamento, agrupados por estado civil.

```
SELECT
    d.department AS Departamento,
    m.maritalstatus AS MaritalStatus,
    ROUND(AVG(p.age), 2) AS MediaIdade
FROM
    personalinfo p
JOIN
    employee e ON p.employeeid = e.employeeid
JOIN
    department d ON e.departmentid = d.departmentid
JOIN
    maritalstatus m ON p.maritalstatusid = m.maritalstatusid
GROUP BY
    d.department,
    m.maritalstatus
ORDER BY
    d.department,
    m.maritalstatus;
```

Excertos dos principais scripts T-SQL

--idade média dos funcionários em cada departamento, de acordo com o estado civil

```
SELECT
CASE
    WHEN p.age BETWEEN 18 AND 25 THEN '18-25'
    WHEN p.age BETWEEN 26 AND 35 THEN '26-35'
    WHEN p.age BETWEEN 36 AND 45 THEN '36-45'
    WHEN p.age BETWEEN 46 AND 55 THEN '46-55'
    ELSE '56+'
END AS AgeGroup,
COUNT(*) AS TotalEmployees,
SUM(CASE WHEN a.attrition = 'Yes' THEN 1 ELSE 0 END) AS TotalAttrition,
FORMAT(
    CAST(SUM(CASE WHEN a.attrition = 'Yes' THEN 1 ELSE 0 END) AS DECIMAL(10,2)) /
    CAST(COUNT(*) AS DECIMAL(10,2)),
    'P0'
) AS GroupAttritionRate
FROM
    employee e
JOIN
    personalinfo p ON e.employeeid = p.employeeid
JOIN
    attrition a ON e.attritionid = a.attritionid
GROUP BY
CASE
    WHEN p.age BETWEEN 18 AND 25 THEN '18-25'
    WHEN p.age BETWEEN 26 AND 35 THEN '26-35'
    WHEN p.age BETWEEN 36 AND 45 THEN '36-45'
    WHEN p.age BETWEEN 46 AND 55 THEN '46-55'
    ELSE '56+'
END
ORDER BY
    MIN(p.age);
```

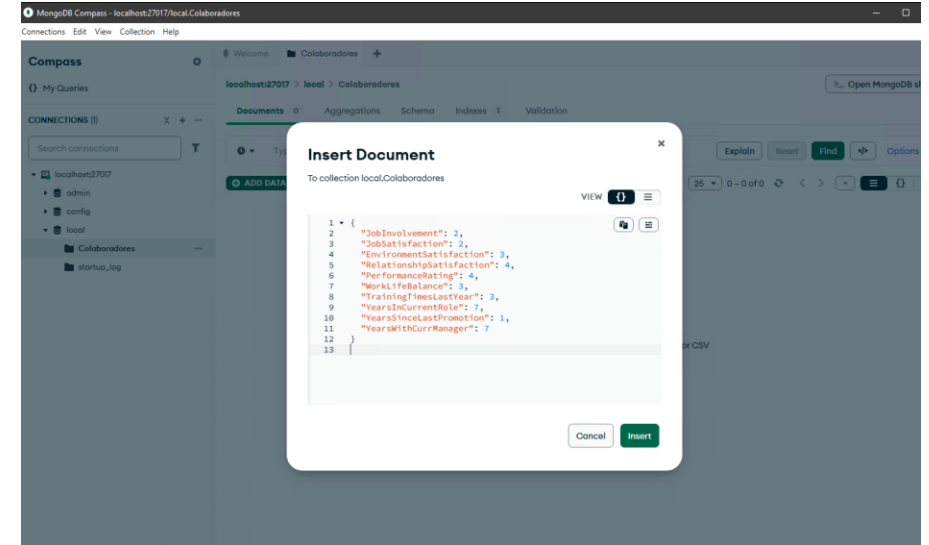

Excertos dos principais scripts T-SQL

```
--relatório completo por departamento, com informações de salários,  
--quantidade de funcionários, número e taxa de rotatividade (attrition).  
SELECT  
    d.department AS Departamento,  
    FORMAT(ROUND(AVG(s.monthlyincome), 2), 'N2', 'pt-PT') + ' €' AS MédiaSalarial,  
    FORMAT(MAX(s.monthlyincome), 'N2', 'pt-PT') + ' €' AS SalárioMáximo,  
    FORMAT(MIN(s.monthlyincome), 'N2', 'pt-PT') + ' €' AS SalárioMínimo,  
    COUNT(*) AS TotalFuncionários,  
    SUM(CASE WHEN a.attrition = 'Yes' THEN 1 ELSE 0 END) AS FuncionáriosAttrition,  
    FORMAT(CAST(SUM(CASE WHEN a.attrition = 'Yes' THEN 1 ELSE 0 END) AS DECIMAL(10,2)) /  
        CAST(COUNT(*) AS DECIMAL(10,2)), 'P0', 'pt-PT') AS TaxaAttrition  
FROM  
    employee e  
JOIN  
    department d ON e.departmentid = d.departmentid  
JOIN  
    attrition a ON e.attritionid = a.attritionid  
JOIN  
    salary s ON e.employeeid = s.employeeid  
GROUP BY  
    d.department  
ORDER BY  
    d.department;
```


Resumo explicado do caso de uso NoSQL

Contexto e Implementação

Para o caso de uso NoSQL, implementamos uma solução utilizando MongoDB para armazenar e analisar dados de desempenho dos colaboradores. O MongoDB foi escolhido devido à sua flexibilidade para armazenar documentos com estruturas que podem evoluir ao longo do tempo, sem necessidade de alterar o esquema. (Documento Suporte Word)

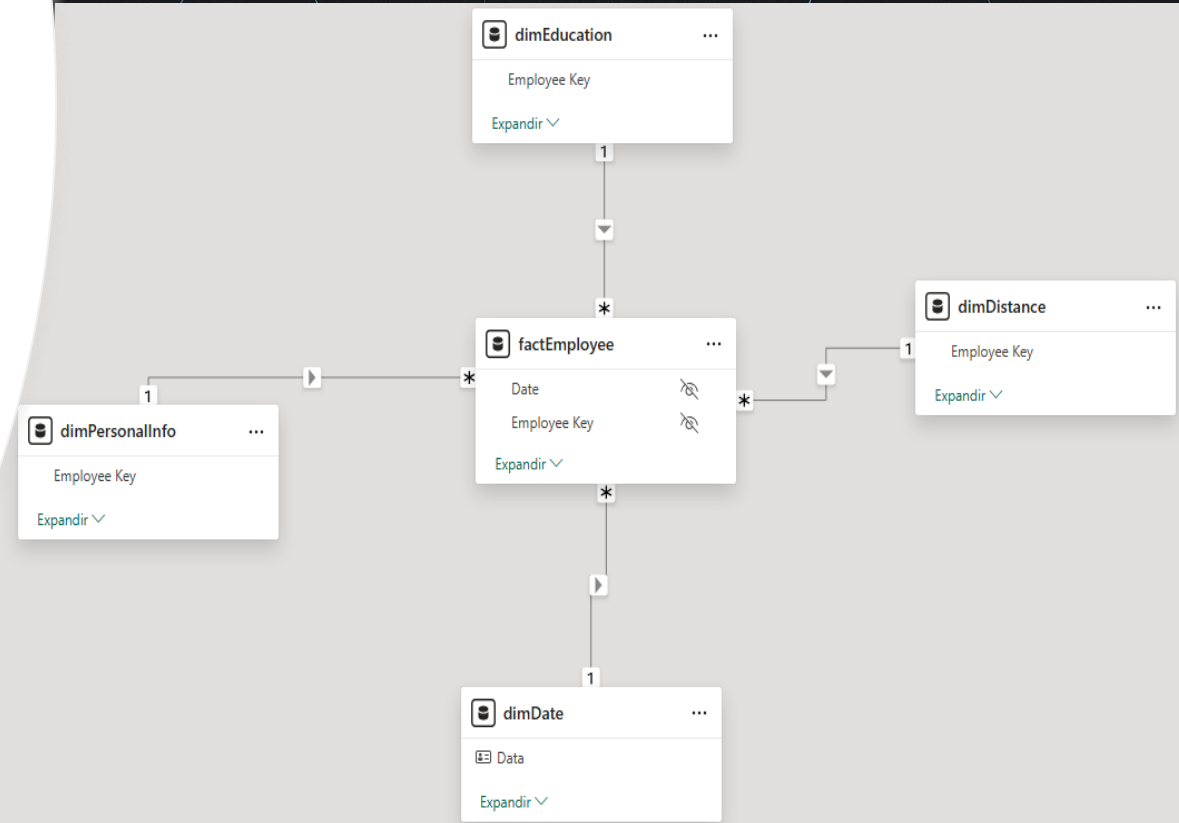


	ifeBalance	TrainingTimesLastYear	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
1	3	7	3	6	
2	3	2	2	2	
3	6	4	0	5	
4	3	7	1	7	
5	3	0	0	0	
6	3	7	3	0	
7	3	2	2	2	
8	2	7	3	6	
9	3	0	0	0	
10	2	0	0	0	
11	2	7	1	8	
12	3	7	7	7	
13	5	4	0	3	
14	3	5	0	8	
15	1	2	4	3	
16	2	2	1	2	
17	4	2	0	3	
18	1	9	0	0	
19	5	2	0	5	
20	2	0	0	0	
21	3	8	3	7	



Comparação visual dos dados antes e depois das transformações

Imagem do modelo OLAP em estrela



Lista das métricas criadas, com a respectiva explicação

Foi criada uma coluna calculada em DAX para realizar uma projeção salarial para 5 anos com base no aumento percentual (*Percent Salary Hike*)

```
c_ProjectedSalary5Years =  
ROUND(  
    factEmployee[Monthly Income] * POWER(1 + (factEmployee[Percent Salary Hike] / 100); 5);  
    2  
)
```

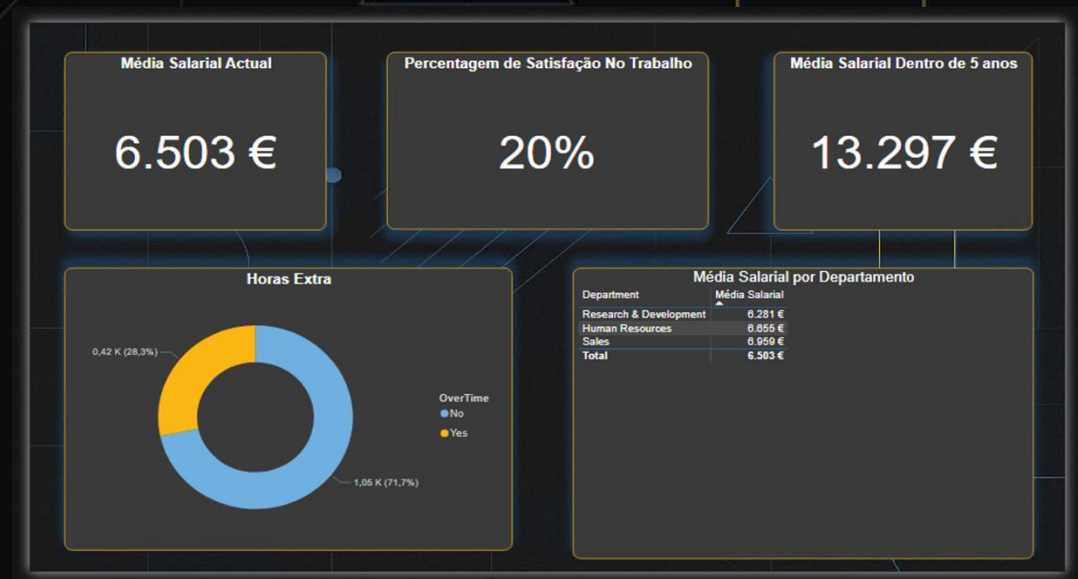
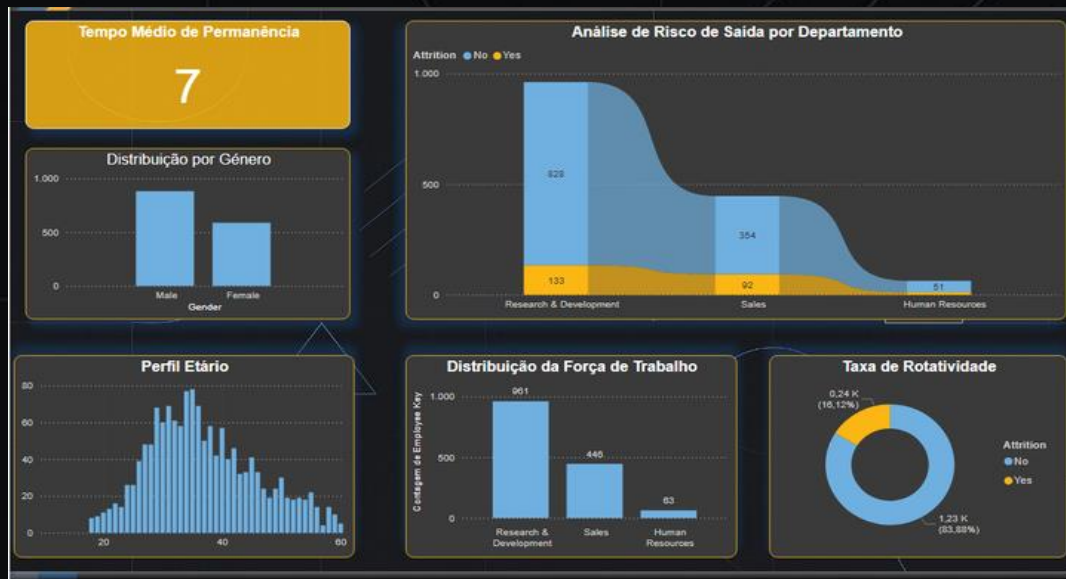
Foi criada uma medida que calcula a taxa de funcionários com baixa satisfação no trabalho.

```
m_LowJobSatisfactionRate =  
DIVIDE(  
    COUNTROWS(FILTER(factEmployee; factEmployee[Job Satisfaction] = "Low"));  
    COUNTROWS(factEmployee);  
    0  
)
```

Foi criada a medida que calcula a média salarial por departamento

```
m_AvgSalaryByDepartment =  
CALCULATE(  
    AVERAGE( factEmployee[Monthly Income] );  
    ALLEXCEPT( factEmployee; factEmployee[Department] )  
)
```


Imagens das principais visualizações



Resumo da análise exploratória dos dados : Análise de Dados de RH - IBM

Contexto

Projeto de análise de dados dos recursos humanos da IBM, realizado por uma equipa multidisciplinar para identificar padrões, tendências e insights relevantes para a gestão de talentos.

Objetivos Alcançados

- Análise completa dos dados de RH da IBM
- Identificação de fatores determinantes para retenção de talentos
- Desenvolvimento de modelos preditivos para rotatividade de pessoal
- Criação de dashboards interativos em Power BI

Principais Descobertas

- Correlações significativas entre satisfação no trabalho e permanência na empresa
- Impacto do equilíbrio trabalho-vida na produtividade dos colaboradores
- Padrões de desenvolvimento de carreira que favorecem a retenção

Resumo da análise exploratória dos dados : Análise de Dados de RH - IBM

Conclusão

O projeto de análise de dados de RH da IBM forneceu insights valiosos sobre os fatores que influenciam a satisfação e retenção dos colaboradores. Os resultados obtidos permitiram identificar áreas críticas para intervenção e desenvolver estratégias baseadas em dados para otimizar a gestão de talentos. A implementação das recomendações propostas tem potencial para reduzir significativamente a rotatividade, aumentar a satisfação dos funcionários e melhorar os indicadores gerais de desempenho organizacional.