# Document HR Analytics SQL queries

**Step 1: Data Preparation**

**Data Import:**

```
-- Create schema fOR HR data
CREATE SCHEMA IF NOT EXISTS hr_analytics;

-- Switch to the new schema
USE hr_analytics;

-- Create the hr_data table
CREATE TABLE hr_data (
 Age INT,
 AttritiON VARCHAR(3),
 BusinessTravel VARCHAR(50),
 DailyRate INT,
 Department VARCHAR(50),
 DistanceFromHome INT,
 EducatiON INT,
 EducationField VARCHAR(50),
 EmployeeCount INT,
 EmployeeNumber INT PRIMARY KEY,
 EnvironmentSatisfactiON INT,
 Gender VARCHAR(10),
 HourlyRate INT,
 JobInvolvement INT,
 JobLevel INT,
 JobRole VARCHAR(50),
 JobSatisfactiON INT,
 MaritalStatus VARCHAR(20),
 MonthlyIncome INT,
 MonthlyRate INT,
 NumCompaniesWorked INT,
 Over18 CHAR(1),
 OverTime VARCHAR(3),
 PercentSalaryHike INT,
 PerformanceRating INT,
 RelationshipSatisfactiON INT,
 StandardHours INT,
 StockOptionLevel INT,
 TotalWorkingYears INT,
 TrainingTimesLastYear INT,
 WorkLifeBalance INT,
```

```
 YearsAtCompany INT,
 YearsInCurrentRole INT,
 YearsSinceLastPromotiON INT,
 YearsWithCurrManager INT
);

-- Load data
    FROM CSV file
LOAD DATA LOCAL INFILE 'C:/ProgramData/MySQL/MySQL Server
8.0/Uploads/hr_data.csv'
INTO TABLE hr_data
FIELDS TERMINATED BY ',
        '
OPTIONALLY ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS
(Age,
        Attrition,
        BusinessTravel,
        DailyRate,
        Department,
        DistanceFromHome,
        Education,
        EducationField,
        EmployeeCount,
        EmployeeNumber,
        EnvironmentSatisfaction,
        Gender,
        HourlyRate,
        JobInvolvement,
        JobLevel,
        JobRole,
        JobSatisfaction,
        MaritalStatus,
        MonthlyIncome,
        MonthlyRate,
        NumCompaniesWorked,
        Over18,
        OverTime,
        PercentSalaryHike,
        PerformanceRating,
        RelationshipSatisfaction,
        StandardHours,
        StockOptionLevel,
        TotalWorkingYears,
        TrainingTimesLastYear,
        WorkLifeBalance,
        YearsAtCompany,
        YearsInCurrentRole,
```

```
        YearsSinceLastPromotion,
        YearsWithCurrManager);

-- Display first few rows to verify data
SELECT *
    FROM hr_data
LIMIT 5;
```

Output:

| Age | Attrition | BusinessTravel | DailyRate | Department | DistanceFromHome | Education | EducationField | EmployeeCount | EmployeeNumber | EnvironmentSatisfaction | Gender | HourlyRate | JobInvo |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 41 | Yes | Travel_Rarely | 1102 | Sales | 1 | 2 | Life Sciences | 1 | 1 | 2 | Female | 94 | 3 |
| 49 | No | Travel_Frequently | 279 | Research & Development | 8 | 1 | Life Sciences | 1 | 2 | 3 | Male | 61 | 2 |
| 37 | Yes | Travel_Rarely | 1373 | Research & Development | 2 | 2 | Other | 1 | 4 | 4 | Male | 92 | 2 |
| 33 | No | Travel_Frequently | 1392 | Research & Development | 3 | 4 | Life Sciences | 1 | 5 | 4 | Female | 56 | 3 |
| 27 | No | Travel_Rarely | 591 | Research & Development | 2 | 1 | Medical | 1 | 7 | 1 | Male | 40 | 3 |
| NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL |

**Data Cleaning:**

```
-- Check fOR missing values IN critical columns (Attrition,
        Business Travel,
        Job Role,
        EducatiON Level)
SELECT COUNT(*) AS missing_attrition
FROM hr_data
WHERE AttritiON IS NULL;

SELECT COUNT(*) AS missing_business_travel
FROM hr_data
WHERE `BusinessTravel` IS NULL;

SELECT COUNT(*) AS missing_job_role
FROM hr_data
WHERE `JobRole` IS NULL;

SELECT COUNT(*) AS missing_education_level
FROM hr_data
WHERE `Education` IS NULL;


-- Replace missing values
WITH default values
UPDATE hr_data
SET AttritiON = 'No'
WHERE AttritiON IS NULL;

UPDATE hr_data
SET `BusinessTravel` = 'NOT Specified'
WHERE `BusinessTravel` IS NULL;
```

```
UPDATE hr_data
SET `JobRole` = 'Unknown'
WHERE `JobRole` IS NULL;

UPDATE hr_data
SET `Education` = 'Unknown'
WHERE `Education` IS NULL;


UPDATE hr_data
SET Age = CAST(Age AS SIGNED),

 MonthlyIncome = CAST(MonthlyIncome AS DECIMAL(10,
         2)),

 YearsAtCompany = CAST(YearsAtCompany AS SIGNED)
WHERE Age IS NOT NULL

        AND MonthlyIncome IS NOT NULL

        AND YearsAtCompany IS NOT NULL;
```

Output:

| Age | Attrition | BusinessTravel | DailyRate | Department | DistanceFromHome | Education | EducationField | EmployeeCount | EmployeeNumber | EnvironmentSatisfaction | Gender | HourlyRate | JobIn |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 41 | Yes | Travel_Rarely | 1102 | Sales | 1 | 2 | Life Sciences | 1 | 1 | 2 | Female | 94 | 3 |
| 49 | No | Travel_Frequently | 279 | Research & Development | 8 | 1 | Life Sciences | 1 | 2 | 3 | Male | 61 | 2 |
| 37 | Yes | Travel_Rarely | 1373 | Research & Development | 2 | 2 | Other | 1 | 4 | 4 | Male | 92 | 2 |
| 33 | No | Travel_Frequently | 1392 | Research & Development | 3 | 4 | Life Sciences | 1 | 5 | 4 | Female | 56 | 3 |

**Data Normalization:**

Code:

```
DELIMITER $$

CREATE PROCEDURE hr_star_schema()
BEGIN

-- Create DimensiON Tables

-- DimensiON fOR Age
CREATE TABLE IF NOT EXISTS dim_age (
 age_id INT AUTO_INCREMENT PRIMARY KEY,

 age_range VARCHAR(255)
);

-- DimensiON fOR AttritiON (Yes/No)
```

```sql
CREATE TABLE IF NOT EXISTS dim_attritiON (
 attrition_id INT AUTO_INCREMENT PRIMARY KEY,

 attrition_status VARCHAR(255)
);

-- DimensiON fOR Business Travel
CREATE TABLE IF NOT EXISTS dim_business_travel (
 business_travel_id INT AUTO_INCREMENT PRIMARY KEY,

 travel_type VARCHAR(255)
);

-- DimensiON fOR Department
CREATE TABLE IF NOT EXISTS dim_department (
 department_id INT AUTO_INCREMENT PRIMARY KEY,

 department_name VARCHAR(255)
);

-- DimensiON fOR Education
CREATE TABLE IF NOT EXISTS dim_educatiON (
 education_id INT AUTO_INCREMENT PRIMARY KEY,

 education_level VARCHAR(255)
);

-- DimensiON fOR Job Role
CREATE TABLE IF NOT EXISTS dim_job_role (
 job_role_id INT AUTO_INCREMENT PRIMARY KEY,

 job_role_name VARCHAR(255)
);

-- DimensiON fOR Gender
CREATE TABLE IF NOT EXISTS dim_gender (
 gender_id INT AUTO_INCREMENT PRIMARY KEY,

 gender_type VARCHAR(255)
);

-- Fact Table fOR HR Employee Data
CREATE TABLE IF NOT EXISTS fact_hr_data (
 employee_number INT PRIMARY KEY,

 age_id INT,

 attrition_id INT,
```

```
business_travel_id INT,
department_id INT,
education_id INT,
job_role_id INT,
gender_id INT,
daily_rate INT,
distance_from_home INT,
environment_satisfactiON INT,
hourly_rate INT,
job_involvement INT,
job_level INT,
job_satisfactiON INT,
marital_status VARCHAR(255),
monthly_income INT,
monthly_rate INT,
num_companies_worked INT,
over18 VARCHAR(255),
over_time VARCHAR(255),
percent_salary_hike INT,
performance_rating INT,
relationship_satisfactiON INT,
standard_hours INT,
stock_option_level INT,
total_working_years INT,
training_times_last_year INT,
```

```
 work_life_balance INT,

 years_at_company INT,

 years_in_current_role INT,

 years_since_last_promotiON INT,

 years_with_curr_manager INT,

 FOREIGN KEY (age_id) REFERENCES dim_age(age_id),

 FOREIGN KEY (attrition_id) REFERENCES dim_attrition(attrition_id),

 FOREIGN KEY (business_travel_id) REFERENCES
dim_business_travel(business_travel_id),

 FOREIGN KEY (department_id) REFERENCES dim_department(department_id),

 FOREIGN KEY (education_id) REFERENCES dim_education(education_id),

 FOREIGN KEY (job_role_id) REFERENCES dim_job_role(job_role_id),

 FOREIGN KEY (gender_id) REFERENCES dim_gender(gender_id)
);

-- Insert Data into DimensiON Tables

-- Insert DISTINCT Age Ranges into dim_age
INSERT INTO dim_age (age_range)
SELECT DISTINCT
 CASE

    WHEN Age
    BETWEEN 18
        AND 24 THEN
    '18-24'

    WHEN Age
    BETWEEN 25
        AND 34 THEN
    '25-34'

    WHEN Age
    BETWEEN 35
        AND 44 THEN
    '35-44'
```

```
    WHEN Age
    BETWEEN 45
        AND 54 THEN
    '45-54'

    WHEN Age
    BETWEEN 55
        AND 64 THEN
    '55-64'

    WHEN Age >= 65 THEN
    '65+'

    END AS age_range
FROM hr_data
WHERE Age IS NOT NULL;

-- Insert DISTINCT AttritiON Status into dim_attrition
INSERT INTO dim_attritiON (attrition_status)
SELECT DISTINCT Attrition
FROM hr_data
WHERE AttritiON IS NOT NULL;

-- Insert DISTINCT Business Travel Types into dim_business_travel
INSERT INTO dim_business_travel (travel_type)
SELECT DISTINCT BusinessTravel
FROM hr_data
WHERE BusinessTravel IS NOT NULL;

-- Insert DISTINCT Departments into dim_department
INSERT INTO dim_department (department_name)
SELECT DISTINCT Department
FROM hr_data
WHERE Department IS NOT NULL;

-- Insert DISTINCT EducatiON Levels into dim_education
INSERT INTO dim_educatiON (education_level)
SELECT DISTINCT Education
FROM hr_data
WHERE EducatiON IS NOT NULL;

-- Insert DISTINCT Job Roles into dim_job_role
INSERT INTO dim_job_role (job_role_name)
SELECT DISTINCT JobRole
FROM hr_data
WHERE JobRole IS NOT NULL;

-- Insert DISTINCT Genders into dim_gender
INSERT INTO dim_gender (gender_type)
```

```sql
SELECT DISTINCT Gender
FROM hr_data
WHERE Gender IS NOT NULL;

-- Insert data into Fact Table fact_hr_data
INSERT INTO fact_hr_data (
 employee_number,
        age_id,
        attrition_id,
        business_travel_id,
        department_id,
        education_id,
        job_role_id,

 gender_id,
        daily_rate,
        distance_from_home,
        environment_satisfaction,
        hourly_rate,
        job_involvement,
        job_level,

 job_satisfaction,
        marital_status,
        monthly_income,
        monthly_rate,
        num_companies_worked,
        over18,
        over_time,

 percent_salary_hike,
        performance_rating,
        relationship_satisfaction,
        standard_hours,
        stock_option_level,

 total_working_years,
        training_times_last_year,
        work_life_balance,
        years_at_company,
        years_in_current_role,

 years_since_last_promotion,
        years_with_curr_manager
)
SELECT
 e.EmployeeNumber,

 a.age_id,
```

```
at.attrition_id,

bt.business_travel_id,

d.department_id,

ed.education_id,

j.job_role_id,

g.gender_id,

e.DailyRate,

e.DistanceFromHome,

e.EnvironmentSatisfaction,

e.HourlyRate,

e.JobInvolvement,

e.JobLevel,

e.JobSatisfaction,

e.MaritalStatus,

e.MonthlyIncome,

e.MonthlyRate,

e.NumCompaniesWorked,

e.Over18,

e.OverTime,

e.PercentSalaryHike,

e.PerformanceRating,

e.RelationshipSatisfaction,

e.StandardHours,

e.StockOptionLevel,
```

```
 e.TotalWorkingYears,

 e.TrainingTimesLastYear,

 e.WorkLifeBalance,

 e.YearsAtCompany,

 e.YearsInCurrentRole,

 e.YearsSinceLastPromotion,

 e.YearsWithCurrManager
FROM hr_data e
LEFT
JOIN dim_age a
    ON CASE

    WHEN e.Age
    BETWEEN 18
        AND 24 THEN
    '18-24'

    WHEN e.Age
    BETWEEN 25
        AND 34 THEN
    '25-34'

    WHEN e.Age
    BETWEEN 35
        AND 44 THEN
    '35-44'

    WHEN e.Age
    BETWEEN 45
        AND 54 THEN
    '45-54'

    WHEN e.Age
    BETWEEN 55
        AND 64 THEN
    '55-64'

    WHEN e.Age >= 65 THEN
    '65+'

    END = a.age_range
LEFT
JOIN dim_attritiON at
```

```
    ON e.AttritiON = at.attrition_status
LEFT
JOIN dim_business_travel bt
    ON e.BusinessTravel = bt.travel_type
LEFT
JOIN dim_department d
    ON e.Department = d.department_name
LEFT
JOIN dim_educatiON ed
    ON e.EducatiON = ed.education_level
LEFT
JOIN dim_job_role j
    ON e.JobRole = j.job_role_name
LEFT
JOIN dim_gender g
    ON e.Gender = g.gender_type;

END $$

DELIMITER ;

-- Call the procedure to normalize the data
CALL hr_star_schema();
```

Code:

```
SELECT *
    FROM hr_analytics.dim_age;
```

Output:

| age_id | age_range |
|---|---|
| 1 | 35-44 |
| 2 | 45-54 |
| 3 | 25-34 |
| 4 | 55-64 |
| 5 | 18-24 |

Code:

```
SELECT *
    FROM hr_analytics.dim_attrition;
```

Output:

| attrition_id | attrition_status |
|---|---|
| 1 | Yes |
| 2 | No |

Code:

```
SELECT
 *
FROM
 Hr_analytics.dim_business_travel;
```

Output:

| business_travel_id | travel_type |
|---|---|
| 1 | Travel_Rarely |
| 2 | Travel_Frequently |
| 3 | Non-Travel |

Code:

```
SELECT *
    FROM hr_analytics.dim_department;
```

Output:

| department_id | department_name |
|---|---|
| 1 | Sales |
| 2 | Research & Development |
| 3 | Human Resources |

Code:

```
SELECT *
    FROM hr_analytics.dim_education;
```

Output:

| education_id | education_level |
|---|---|
| 1 | 2 |
| 2 | 1 |
| 3 | 4 |
| 4 | 3 |
| 5 | 5 |

Code:

```
SELECT *
```

```
    FROM hr_analytics.dim_job_role;
```

Output:

| job_role_id | job_role_name |
|---|---|
| 1 | Sales Executive |
| 2 | Research Scientist |
| 3 | Laboratory Technician |
| 4 | Manufacturing Director |
| 5 | Healthcare Representative |
| 6 | Manager |
| 7 | Sales Representative |
| 8 | Research Director |
| 9 | Human Resources |

Code:

```
SELECT *
    FROM hr_analytics.fact_hr_data;
```

Output:

| employee_number | age_id | attrition_id | business_travel_id | department_id | education_id | job_role_id | gender_id | daily_rate | distance_from_home | environment_satisfaction | hourly_rate | job_involvement | jo |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1102 | 1 | 2 | 94 | 3 | 2 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 279 | 8 | 3 | 61 | 2 | 2 |
| 4 | 1 | 1 | 1 | 2 | 1 | 3 | 2 | 1373 | 2 | 4 | 92 | 2 | 1 |
| 5 | 3 | 2 | 2 | 2 | 3 | 2 | 1 | 1392 | 3 | 4 | 56 | 3 | 1 |
| 7 | 3 | 2 | 1 | 2 | 2 | 3 | 2 | 591 | 2 | 1 | 40 | 3 | 1 |
| 8 | 3 | 2 | 2 | 2 | 1 | 3 | 2 | 1005 | 2 | 4 | 79 | 3 | 1 |
| 10 | 4 | 2 | 1 | 2 | 4 | 3 | 1 | 1324 | 3 | 3 | 81 | 4 | 1 |