

```
USE sf;

SELECT * FROM Salaries;

-- data cleaning

-- datatypes

SELECT COLUMN_NAME, DATA_TYPE FROM INFORMATION_SCHEMA.COLUMNS

-- Backup your table/data before making changes
-- Step 1: Identify non-convertible data
SELECT BasePay
FROM Salaries
WHERE ISNUMERIC(BasePay) = 0;

-- Step 2: Replace non-convertible data with '0'
UPDATE Salaries
SET BasePay = '0'
WHERE BasePay = 'Not Provided';

-- Step 3: Convert to float
ALTER TABLE Salaries
ALTER COLUMN BasePay FLOAT;
SELECT COLUMN_NAME, DATA_TYPE FROM INFORMATION_SCHEMA.COLUMNS

-- overtimepay column

-- Backup your table/data before making changes
-- Step 1: Identify non-convertible data
SELECT OvertimePay
FROM Salaries
WHERE ISNUMERIC(BasePay) = 0;

-- Step 2: Replace non-convertible data with '0'
UPDATE Salaries
SET OvertimePay = '0'
WHERE OvertimePay = 'Not Provided';

-- Step 3: Convert to float
ALTER TABLE Salaries
ALTER COLUMN OvertimePay FLOAT;
SELECT COLUMN_NAME, DATA_TYPE FROM INFORMATION_SCHEMA.COLUMNS

-- otherpay

-- Backup your table/data before making changes
-- Step 1: Identify non-convertible data
SELECT OtherPay
FROM Salaries
```

---

```
WHERE ISNUMERIC(BasePay) = 0;
```

```
-- Step 2: Replace non-convertible data with '0'
```

```
UPDATE Salaries
```

```
SET OtherPay = '0'
```

```
WHERE OtherPay = 'Not Provided';
```

```
-- Step 3: Convert to float
```

```
ALTER TABLE Salaries
```

```
ALTER COLUMN OtherPay FLOAT;
```

```
SELECT COLUMN_NAME, DATA_TYPE FROM INFORMATION_SCHEMA.COLUMNS
```

```
--total pay
```

```
-- Step 1: Identify non-convertible data
```

```
SELECT TotalPay
```

```
FROM Salaries
```

```
WHERE ISNUMERIC(BasePay) = 0;
```

```
-- Step 2: Replace non-convertible data with '0'
```

```
UPDATE Salaries
```

```
SET TotalPay = '0'
```

```
WHERE TotalPay = 'Not Provided';
```

```
-- Step 3: Convert to float
```

```
ALTER TABLE Salaries
```

```
ALTER COLUMN TotalPay FLOAT;
```

```
SELECT COLUMN_NAME, DATA_TYPE FROM INFORMATION_SCHEMA.COLUMNS
```

```
--total pay benefits
```

```
-- Step 1: Identify non-convertible data
```

```
SELECT TotalPayBenefits
```

```
FROM Salaries
```

```
WHERE ISNUMERIC(BasePay) = 0;
```

```
-- Step 2: Replace non-convertible data with '0'
```

```
UPDATE Salaries
```

```
SET TotalPayBenefits = '0'
```

```
WHERE TotalPayBenefits = 'Not Provided';
```

```
-- Step 3: Convert to float
```

```
ALTER TABLE Salaries
```

```
ALTER COLUMN TotalPayBenefits FLOAT;
```

```
SELECT COLUMN_NAME, DATA_TYPE FROM INFORMATION_SCHEMA.COLUMNS
```

```
--Year
```

```
-- Step 1: Identify non-convertible data
```

```
SELECT Year
FROM Salaries
WHERE ISNUMERIC(BasePay) = 0;
```

```
-- Step 2: Replace non-convertible data with '0'
```

```
UPDATE Salaries
SET Year = '0'
WHERE Year = 'Not Provided';
```

```
-- Step 3: Convert to float
```

```
ALTER TABLE Salaries
ALTER COLUMN Year INT;
SELECT COLUMN_NAME, DATA_TYPE FROM INFORMATION_SCHEMA.COLUMNS
```

```
-----
-- how many null value
```

```
SELECT
    SUM(CASE WHEN BasePay IS NULL THEN 1 ELSE 0 END) AS BasePay_NullCount,
    SUM(CASE WHEN OverTimePay IS NULL THEN 1 ELSE 0 END) AS OverTimePay_NullCount,
    SUM(CASE WHEN OtherPay IS NULL THEN 1 ELSE 0 END) AS OtherPay_NullCount,
    SUM(CASE WHEN Benefits IS NULL THEN 1 ELSE 0 END) AS Benefits_NullCount,
    SUM(CASE WHEN OverTimePay IS NULL THEN 1 ELSE 0 END) AS OverTimePay_NullCount,
    SUM(CASE WHEN TotalPay IS NULL THEN 1 ELSE 0 END) AS TotalPay_NullCount,
    SUM(CASE WHEN TotalPayBenefits IS NULL THEN 1 ELSE 0 END) AS
        TotalPayBenefits_NullCount,
    SUM(CASE WHEN Year IS NULL THEN 1 ELSE 0 END) AS Year_NullCount,
    -- SUM(CASE WHEN Notes IS NULL THEN 1 ELSE 0 END) AS Notes_NullCount,
    SUM(CASE WHEN Agency IS NULL THEN 1 ELSE 0 END) AS Agency_NullCount,
    SUM(CASE WHEN Status IS NULL THEN 1 ELSE 0 END) AS Status_NullCount
FROM
    Salaries;
```

```
-- delete notes column
```

```
ALTER TABLE Salaries
DROP COLUMN Notes;
```

```
-- replace nan values in basepay from nan to zero
```

```
UPDATE Salaries
SET BasePay = 0
WHERE BasePay IS NULL;
```

```
-- replace nan values in benefits from nan to zero
```

```
UPDATE Salaries
SET Benefits = 0
WHERE Benefits IS NULL;
```

```
-- replace nan values in status from nan to zero
```

```
UPDATE Salaries
SET Status = 'unknown'
WHERE Status IS NULL;

-- what is average basepay
SELECT AVG(BasePay) AS AverageBasePay
FROM Salaries;

-- what is the highest amount of overtimepay in dataset
SELECT MAX(OvertimePay) AS MaxOvertimePay
FROM Salaries;

-- what is job title of joseph driscoll
SELECT EmployeeName, JobTitle
FROM Salaries
WHERE EmployeeName = 'JOSEPH DRISCOLL';

--how much does joseph driscoll make include benefits
SELECT EmployeeName, TotalPayBenefits
FROM Salaries
WHERE TotalPayBenefits = (
    SELECT MAX(TotalPayBenefits)
    FROM Salaries
);

-- What is the name of the person with the lowest total pay and benefits
SELECT EmployeeName, TotalPayBenefits
FROM Salaries
WHERE TotalPayBenefits = (
    SELECT MIN(TotalPayBenefits)
    FROM Salaries
);

--what was the average BasePay of all employees per year? (2011-2014) ?
SELECT Year, AVG(BasePay) AS AverageBasePay
FROM Salaries
GROUP BY Year;

-- what are the top 5 most highest paid jobs ?
SELECT TOP 5 JobTitle, TotalPayBenefits
FROM Salaries
ORDER BY TotalPayBenefits DESC;
```