

# SQL ANALYSIS PROJECT

Analyzing Departmental Salary Disparities



# Business Case

The project focuses on analyzing salary patterns across departments to identify disparities. The objective is to uncover departments with salaries either above or below their respective departmental averages. This analysis aims to provide actionable insights for addressing salary inconsistencies, ensuring fairness, and optimizing compensation strategies within the organization.

# Project Scope and Key Elements

## Objective

Create a query that identifies a high amount of variation in departmental salaries

## Deliverable

List from a SQL database with Average Salary, department, and a way to score variation.

## Hypothesis

PWD Department has been flagged as a department that has a high amount of salary spread.

# SQL CONCEPTS

- Common Table Expression (CTE)
- CASE Statement
- JOIN
- Aggregation Functions
- Outliers, Standard deviation, Variation
- WHERE clause
- ORDER BY, GROUP BY



# Query Breakdown

Query 1: Calculate the statistics by department

Query 2: Create outlier table

Query 3: Filter and sort the result by variation score



# SQL Query

This query calculates the average salary, standard deviation, coefficient of variation, and outlier count for each department, identifying salary discrepancies.

```
WITH Department_Stat AS
(SELECT Department,
     STDEV(Salary) AS Standard_Deviation,
     AVG(salary) AS Average
  FROM salaries
 WHERE Salary >=10000
 GROUP BY Department
),
Department_Outliers AS(
SELECT emp.Department, emp.Salary,
       (emp.Salary - dp.Average) / dp.Standard_Deviation AS z_score
  FROM salaries Emp
 JOIN Department_Stat dp ON Emp.Department = dp.Department
 WHERE emp.Salary > 10000)
SELECT dp.Department, ROUND(dp.Average,2) AS Average, ROUND(dp.Standard_Deviation,2) AS Std_dev,
       ROUND((dp.Standard_Deviation/dp.Average),2)*100 AS coefficient_of_variation,
       SUM(CASE WHEN(do.z_score >1.96 or do.z_score <-1.96)THEN 1 ELSE 0 END) AS outlier_count
  FROM Department_Stat dp
 LEFT JOIN Department_Outliers as do ON dp.Department = do.Department
 GROUP BY dp.Department, dp.Average, dp.Standard_Deviation, dp.Standard_Deviation/dp.Average
 ORDER BY coefficient_of_variation DESC, dp.Standard_Deviation DESC, outlier_count DESC;
```

# SQL Query - Result

	Department	Average	Std_dev	coefficient_of_variation	outlier_count
1	CMD	107711.81	68918.53	64	1
2	AGR	63457.62	34426.74	54	1
3	CCC	47720.64	25641.94	54	2
4	GRD	53777.7	27133.11	50	1
5	CAD	93902.41	45415.42	48	1
6	CWA	77204.04	36869.17	48	3
7	CUL	63193.65	30134.15	48	1
8	MUS	48815.84	22527.68	46	6
9	CLK	59403.76	25562.12	43	1
10	COR	56679.49	24278.21	43	2
11	MSB	71015.97	30136.39	42	1
12	CIR	64061.38	27153.87	42	1
13	PWD	54081.39	22379.82	41	47
14	AUD	88504.13	35562.14	40	0
15	STR	76835.76	31000.12	40	0

# Top 5 Departments with salary discrepancies

Based on the metrics (CoV, outliers, and standard deviation), the following departments exhibit the highest levels of salary variance and discrepancies.

01. **CMD**

High CV, high standard deviation and some outliers

02. **CAD**

High CV, fairly high standard deviation and some outliers

03. **CCC**

High CV, fair numbers of outliers

04. **PWC**

High number of outliers and fair CV.

05. **CWA**

Fairly high standard deviation, high CV and fair numbers of outliers