KRYSTAL CEBOLLERO: HR INDEPENDENT PROJECT

This is an independent project that I have been working on.

If interested in more of my work, please check out my portfolio at https://www.datacamp.com/portfolio/kwheeler5250.

HR Independent Project contains 2 datasets: employees and performance_schema. Since we are looking more within the realm of HR, we will be reviewing the employees dataset.

The data will be reviewed to see if there is a discrpency in gender bias within the company. This will included determining if there is gender bias within the number of employees within the company and each department. Also, we will be determining if there are pay discrepencies amongst the different genders and departments. (The data will only be containing current employees at this time.)

The employees dataset contains 6 tables.

1. departments
2. dept_emp
3. dept_manager
4. employees
5. salaries
6. titles

To begin let's take a look at the different types of departments that are within the company.

Employees DataFrame as df6

```
SELECT *
FROM departments
ORDER BY dept_no;
```

index	dept_no	dept_name
0	d001	Marketing
1	d002	Finance
2	d003	Human Resources
3	d004	Production
4	d005	Development
5	d006	Quality Management
6	d007	Sales
7	d008	Research
8	d009	Customer Service

Rows: 9 Expand

As shown, there are 9 departments within the company.

Now let's take a look at the employees within each department.

First we will want ot know how many employees are currently working for the company. Then we will want to now how many employees work for each department. We will be wanting to see current employees which have been labeled in the to_date column as '9999-01-01T00:00:00.000Z'.

Employees DataFrame as df8

```
SELECT COUNT(DISTINCT emp_no) AS 'employee count'
FROM dept_emp
WHERE to_date LIKE '9999%';
```

index	employee count
0	

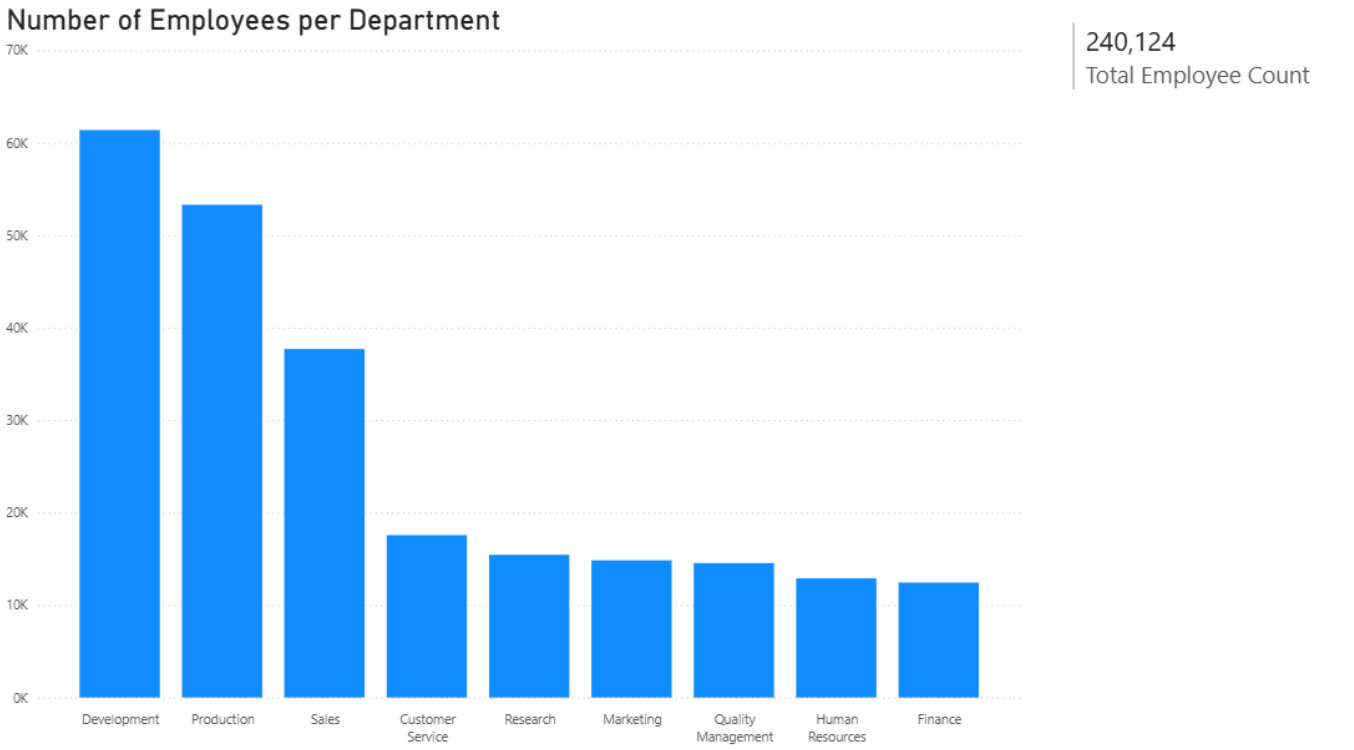
Rows: 1 Expand

Employees DataFrame as df7

```
SELECT departments.dept_name AS 'department name',
    COUNT(*) AS 'employee count'
FROM dept_emp
JOIN departments ON
    dept_emp.dept_no = departments.dept_no
WHERE to_date LIKE '9999%'
GROUP BY dept_name
ORDER BY `employee count` DESC;
```

index	department name	employee count
0	Development	
1	Production	
2	Sales	
3	Customer Service	
4	Research	
5	Marketing	
6	Quality Management	
7	Human Resources	
8	Finance	

Rows: 9 Expand



As you can see, the top three employed departments are: Development, Production, and Sales.

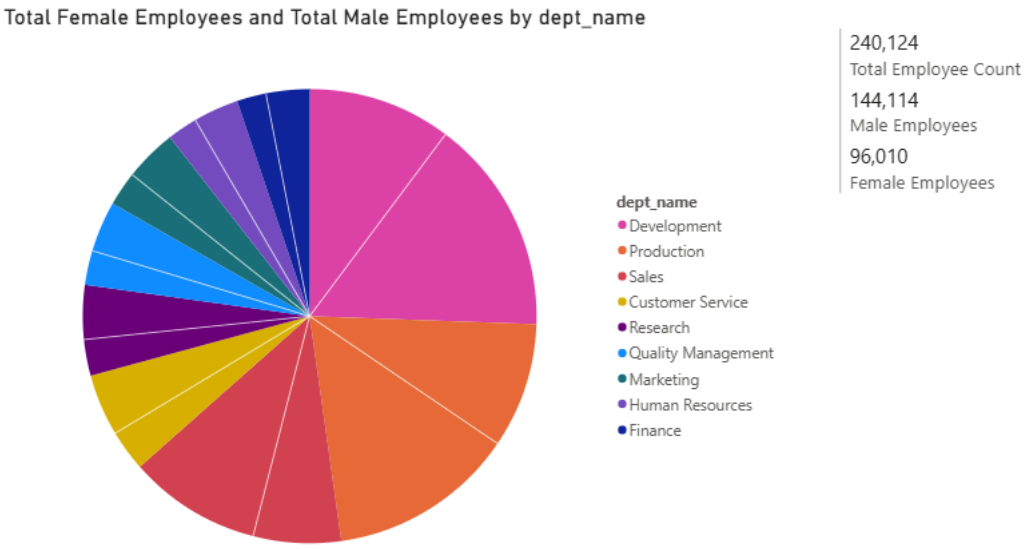
Let's take a closer look at how many males and females are within each department. This is important to see the gender culture within the company.

Employees DataFrame as

```
SELECT departments.dept_name AS department_name,
    COUNT(CASE WHEN employees.gender = 'M' THEN 1 END) AS 'Male employees',
    COUNT(CASE WHEN employees.gender = 'F' THEN 1 END) AS 'Female employeest',
    COUNT(*) AS 'Total employees in department'
FROM dept_emp
JOIN departments ON
    dept_emp.dept_no = departments.dept_no
JOIN employees ON
    dept_emp.emp_no = employees.emp_no
WHERE to_date LIKE '9999%'
GROUP BY department_name
ORDER BY `Total employees in department` DESC;
```

	department_na...	Male em...	Female empl...	Total employees in department
0	Development	36853	24533	61386
1	Production	31911	21393	53304
2	Sales	22702	14999	37701
3	Customer Service	10562	7007	17569
4	Research	9260	6181	15441
5	Marketing	8978	5864	14842
6	Quality Management	8674	5872	14546
7	Human Resources	7751	5147	12898
8	Finance	7423	5014	12437

Rows: 9 Expand



The data indicates that there are more males working within each department. However, let's see how the numbers compare to percentages.

Employees DataFrame as

```
WITH s AS (
    SELECT
        departments.dept_name AS department_name,
        COUNT(CASE WHEN employees.gender = 'M' THEN 1 END) AS `Male employees`,
        COUNT(CASE WHEN employees.gender = 'F' THEN 1 END) AS `Female employees`,
        COUNT(*) AS `Total employees in department`
    FROM employees.dept_emp
    JOIN employees.departments ON
        dept_emp.dept_no = departments.dept_no
    JOIN employees.employees ON
        dept_emp.emp_no = employees.emp_no
    WHERE dept_emp.to_date LIKE '9999%'
    GROUP BY departments.dept_name
)

SELECT
    department_name,
    (`Male employees` / `Total employees in department`) * 100 AS `Percent Male Employees`,
    (`Female employees` / `Total employees in department`) * 100 AS `Percent Female Employees`,
    (`Male employees` / `Total employees in department`) * 100 -
    (`Female employees` / `Total employees in department`) * 100 AS `Difference in Percentages`,
    `Total employees in department`
FROM s
ORDER BY `Total employees in department` DESC;
```

	department_na...	Percent Male Employe...	Percent Female Employees	Difference in Percentages	Total employe
0	Development	60.0349	39.9651	20.0697	
1	Production	59.8661	40.1339	19.7321	
2	Sales	60.2159	39.7841	20.4318	
3	Customer Service	60.1173	39.8827	20.2345	
4	Research	59.9702	40.0298	19.9404	
5	Marketing	60.4905	39.5095	20.981	
6	Quality Management	59.6315	40.3685	19.263	
7	Human Resources	60.0946	39.9054	20.1892	
8	Finance	59.6848	40.3152	19.3696	

Rows: 9 Expand

The data shows that each department has about 20% difference between male versus female employees with each department being male dominant. The percentage of males versus females for each departmert can be rounded to 60% males and 40% females. Due to this, a visual of the percentage of male versus female employees per department is not warranted.

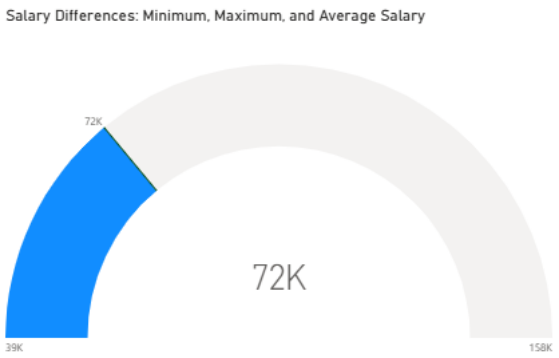
Based on these findings, let's start by determining what id the highest and lowest salaries offered within the company and what the average salary is.

Employees DataFrame as

```
SELECT MAX(salary),
    MIN(salary),
    ROUND(AVG(salary),0) AS 'avg_salary'
FROM salaries
WHERE to_date LIKE '9999%';
```

	MA...	MI...	a...
0	158220	38623	72012

Rows: 1 Expand



Now that we can see what the maximum, minimum, and average salaries are. Let's see how the salaries differ by gender.

Due to there being a difference in the number of employees based on gender, it is important to see if there is a difference in pay for each gender as this could e considered gender discrimination with pay gaps.

Employees DataFrame as

```
SELECT e.gender,
    MAX(s.salary) AS 'max_salary',
    MIN(s.salary) AS 'min_salary',
    ROUND(AVG(s.salary),0) AS 'avg_salary'
FROM salaries AS s
JOIN employees AS e
    ON s.emp_no = e.emp_no
WHERE to_date LIKE '9999%'
GROUP BY e.gender;
```

		m...	m...	a...
0	M	158220	38623	72045
1	F	152710	38936	71964

Rows: 2 Expand

Based on the data, males and females are given similar salaries.

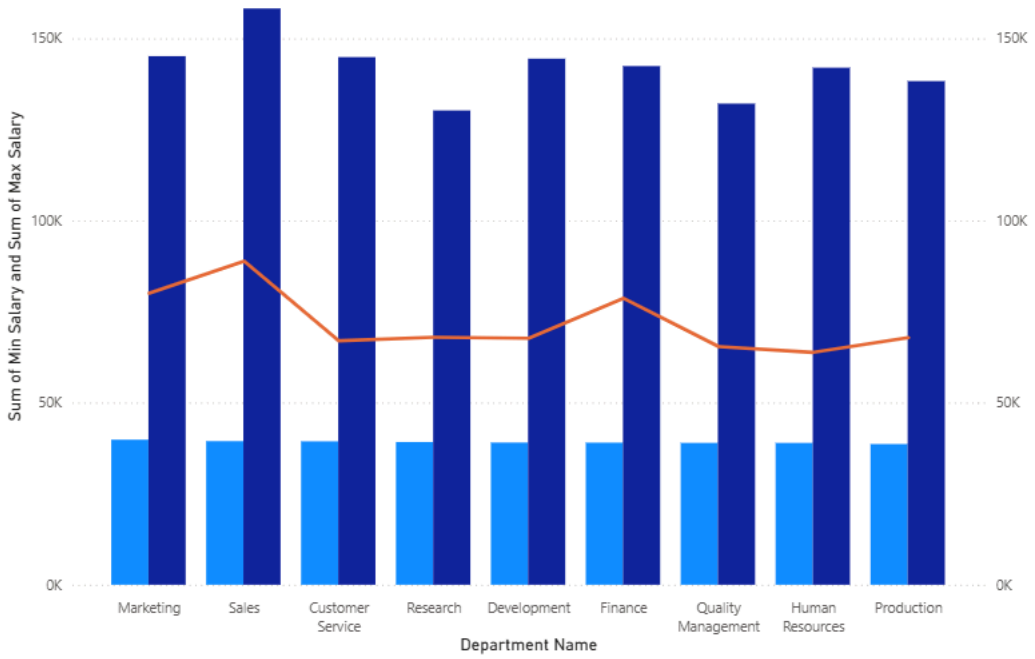
Now that we see that gender discrimination is not an issue within the company. Let's take a look to see how the departments range between maximum, minimum, and average salaries. This can determine a general guideline for companies to provide potential new employees and job websites.

Employees DataFrame as

```
SELECT dept.dept_name AS 'dept_name',
    MAX(s.salary) AS 'max_salary',
    MIN(s.salary) AS 'min_salary',
    ROUND(AVG(s.salary),0) AS 'avg_salary'
FROM salaries AS s
JOIN dept_emp AS emp
    ON s.emp_no = emp.emp_no
JOIN departments AS dept
    ON emp.dept_no = dept.dept_no
WHERE s.to_date LIKE '9999%'
GROUP BY dept.dept_name;
```

	dept_name	m...	m...	a...
0	Development	144434	39036	67666
1	Sales	158220	39426	88842
2	Production	138273	38623	67842
3	Human Resources	141953	38936	63795
4	Research	130211	39186	67933
5	Quality Management	132103	38942	65382
6	Marketing	145128	39821	80015
7	Customer Service	144866	39373	66971
8	Finance	142395	39012	78645

Rows: 9 Expand



The data indicates that all departments contain employees that have salaries over 120K and and minimum salaries around 38-39K. The minimum salary for each department is similar. This could represent a base salary for either the company or for each department to be similar. However, the average salary is where the company can see how pay differs. It appears that sales has the highest average salary near 90K whereas Human Resources has the lowest average salary near 64K.

Now that we know the highest salary given is above 150K and the lowest salary is below 40K, we can take a look at the different types of salaries that are given. The data indicates that the average salary for employees is about 72K. Based on this, it is assumed that most employees would be within a pay bracket containig 72K. Let's determine if that assumption is correct by breaking the data into salary subcategories.

For this inquiry, we will be breaking the salaries down into 5 groups: -Above 150K -Between 120K and 150K -Between 90K and 120K -Between 60K and 90K -Below 60K

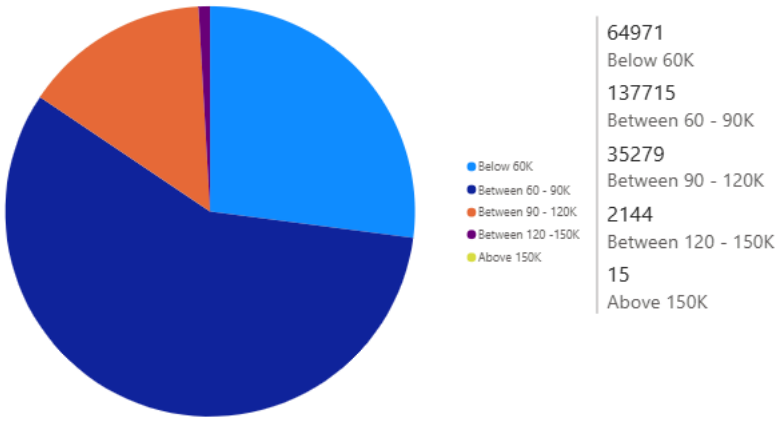
Employees DataFrame as

```
SELECT
    COUNT(CASE WHEN s.salary >= 150000 THEN 1 END) AS 'Above $150K',
    COUNT(CASE WHEN s.salary BETWEEN 120000 AND 149999 THEN 1 END) AS 'Between $120K - $150K',
    COUNT(CASE WHEN s.salary BETWEEN 90000 AND 119999 THEN 1 END) AS 'Between $90K - $120K',
    COUNT(CASE WHEN s.salary BETWEEN 60000 AND 89999 THEN 1 END) AS 'Between $60K - $90K',
    COUNT(CASE WHEN s.salary < 60000 THEN 1 END) AS 'Below 60K'
FROM employees.salaries AS s
    INNER JOIN (
        SELECT emp_no,
        MAX(from_date) AS max_from_date
        FROM employees.salaries
        GROUP BY emp_no
    ) AS n
    ON s.emp_no = n.emp_no AND s.from_date = n.max_from_date
WHERE s.to_date LIKE '9999%';
```

	Ab...	Between $120K - $15...	Between $90K - $1...	Between $60K - ...	B
0	15	2144	35279	137715	64971

Rows: 1 Expand

Total Current Employee Salary Groups



The majority of salaries are between 60K and 90K. This means that the assumption of the most employees being within the bracket containing 72K was correct.

Since we looked at gender and department statistics previously, let's take a deeper look into how salaries differ between the different types of genders and departments.

Employees DataFrame as

```
WITH gndr AS (
    SELECT
        s.emp_no,
        COUNT(CASE WHEN s.salary >= 150000 THEN 1 END) AS above_150,
        COUNT(CASE WHEN s.salary BETWEEN 120000 AND 149999 THEN 1 END) AS between_120_and_150,
        COUNT(CASE WHEN s.salary BETWEEN 90000 AND 119999 THEN 1 END) AS between_90_and_120,
        COUNT(CASE WHEN s.salary BETWEEN 60000 AND 89999 THEN 1 END) AS between_60_and_90,
        COUNT(CASE WHEN s.salary < 60000 THEN 1 END) AS below_60,
        ROUND(AVG(s.salary), 0) AS avg_salary
    FROM employees.salaries as s
        INNER JOIN (
            SELECT emp_no,
            MAX(from_date) AS max_from_date
            FROM employees.salaries
            GROUP BY emp_no
        ) AS n
        ON s.emp_no = n.emp_no AND s.from_date = n.max_from_date
    WHERE s.to_date LIKE '9999%'
    GROUP BY s.emp_no
)

SELECT
    d2.dept_name,
    e.gender,
    SUM(gndr.above_150) AS`Above 150K`,
    SUM(gndr.between_120_and_150) AS `Between 120K and 150K`,
    SUM(gndr.between_90_and_120) AS `Between 90K and 120K`,
    SUM(gndr.between_60_and_90) AS `Between 60 K and 90K`,
    SUM(gndr.below_60) AS `Below 60K`,
    gndr.avg_salary AS `Avg Salary`
FROM gndr
JOIN employees.employees AS e
    ON gndr.emp_no = e.emp_no
JOIN employees.dept_emp AS d1
    ON e.emp_no = d1.emp_no
JOIN employees.departments AS d2
    ON d1.dept_no = d2.dept_no
WHERE d1.to_date LIKE '9999%'
GROUP BY d2.dept_name, e.gender
ORDER BY dept_name;
```

	dept_name		A...	Between 120K and 1...	Between 90K and 1...	Between 60 K and ...	B
0	Customer Service	M	0	63	912	5735	
1	Customer Service	F	0	42	615	3829	
2	Development	M	0	40	2706	22062	
3	Development	F	0	18	1716	14769	
4	Finance	M	0	83	1766	4436	
5	Finance	F	0	55	1190	3061	
6	Human Resources	M	0	6	261	4223	
7	Human Resources	F	0	1	204	2802	
8	Marketing	M	0	121	2527	5165	
9	Marketing	F	0	77	1547	3474	
10	Production	M	0	25	2405	19250	
11	Production	F	0	14	1510	12991	
12	Quality Management	M	0	4	432	4955	
13	Quality Management	F	0	1	273	3420	
14	Research	M	0	12	691	5599	
15	Research	F	0	7	474	3766	

Rows: 18 Expand

A dashboard can be created for the above data to show either/ both the differences in salary based on each department and/or gender based on the company's preference.

Based on the data, sales is the only department where an employees earns more than 150K.

Women have a higher average salary within Finanace, Marketing, Production, Quality Management, and Research. Whereas men have a higher average salary within Customer Service, Development, Human Resources, and Sales. However, both men and women have similar average salaries within the Sales department.

This information is important for determining pay equivolency for new hire and current employees when pay raises are issued. Looking at salary data is important for determining salaries for current employees and new hires. This information is also beneficial for the company is a law suit or a grievance was ever filed within a workers union or against the company for fair pay.