

## **HUMAN RESOURCES – ANALYSIS/CORRELATION**

**PROJECT PURPOSE:** *To analyze the HR dataset to review factors related to recruitment/retention efforts and correlation between employee satisfaction, engagement, and performance.*

### **Objectives:**

- Determine where recruitment efforts can be improved
- Identify the most prominent factors that lead to voluntarily employee departure as a focus to improve employee retention
- Determine the correlation between (if exists):
  - Employee Satisfaction and Engagement
  - Employee Satisfaction and Performance
  - Engagement and Performance

```

USE HumanResources;

/* ----- */
/* ----- RECRUITMENT ANALYSIS/CORRELATION ----- */
/* ----- */

-- CREATE COPY OF DATASET --

SELECT * INTO HRdatacopy
FROM
    (SELECT * FROM [HRDataset_v14(clean)]) AS HRdatacopy;

SELECT * FROM HRdatacopy;

-- CHECK FOR DUPLICATES ROWS IN HRdatacopy DATA --

WITH duplicates AS (
    SELECT Employee_Name,
           EmpID,
           Salary,
           DOB,
           ManagerID,
           ROW_NUMBER() OVER(PARTITION BY Employee_Name, EmpID, Salary, DOB,
                                ManagerID
                                ORDER BY Employee_Name, EmpID, Salary, DOB, ManagerID) AS ROW_NUM
    FROM HRdatacopy)
SELECT * FROM duplicates WHERE ROW_NUM > 1;

```

### **Exploratory Data Analysis (EDA)**

/\* BASIC DEMOGRAPHICS \*/

#### **Female vs. Male Count**

-- NUMBER OF FEMALES/MALES --

```

SELECT Sex,
       COUNT(Sex) AS 'Count'
FROM HRdatacopy
GROUP BY Sex;

```

Sex	Count
F	176
M	135

### Distribution of Race Description (accounting for Hispanic/Latino distinction)

-- DISTRIBUTION OF RACE DESCRIPTION (WHILE ALSO ACCOUNTING FOR HISPANIC/LATINO DISTINCTION) --

```
SELECT RaceDesc,  
       COUNT(RaceDesc) AS 'Count',  
       COUNT(CASE  
           WHEN HispanicLatino = 'Yes' THEN 1  
           ELSE NULL  
       END) AS 'Hispanic/Latino'  
FROM HRdatacopy  
GROUP BY RaceDesc  
ORDER BY 'Count' DESC;
```

RaceDesc	Count	Hispanic/Latino
White	187	18
Black or African American	80	8
Asian	29	0
Two or more races	11	1
American Indian or Alaska Native	3	0
Hispanic	1	1

### Age Distribution

-- AGE DISTRIBUTION --

```
WITH age AS (  
    SELECT DATEDIFF(YEAR, DOB, CONVERT(VARCHAR, GETDATE(), 1)) AS 'Age'  
    FROM HRdatacopy)  
SELECT *, COUNT(Age) AS 'Count'  
FROM age  
GROUP BY age;
```

Age	Count
30	3
31	3
32	5
33	14
34	18
35	19
36	21
37	13
38	16
39	21
40	8

41	14
42	9
43	19
44	8
45	11
46	9
47	9
48	12
49	9
50	7
51	2
52	12
53	7
54	7
55	4
56	4
57	5
58	5
59	3
61	1
63	1
64	2
67	2
68	2
69	1
70	3
71	2

## Citizenship

-- CITIZENSHIP --

```
SELECT CitizenDesc,
       COUNT(CitizenDesc) AS 'Count'
FROM HRdatacopy
GROUP BY CitizenDesc
ORDER BY 'Count' DESC;
```

CitizenDesc	Count
US Citizen	295
Eligible NonCitizen	12
Non-Citizen	4

```
/* BASIC HR INFORMATION/DISTRIBUTION */
```

### Active vs. Inactive Employees

```
-- DISTRIBUTION OF EMPLOYEES IN DATASET (ACTIVE/NON-ACTIVE) --
```

```
SELECT  
COUNT(CASE  
    WHEN EmploymentStatus = 'Active' THEN 1  
    ELSE NULL  
END) AS 'Active',  
COUNT(CASE  
    WHEN EmploymentStatus = 'Terminated for Cause' OR  
    EmploymentStatus = 'Voluntarily Terminated' THEN 1  
    ELSE NULL  
END) AS 'Non-active'  
FROM HRdatacopy;
```

Active	Non-active
207	104

```
-- NUMBER OF DEPARTMENTS --
```

```
SELECT COUNT(DISTINCT Department) AS 'Number of Departments'  
FROM HRdatacopy
```

### List of Departments

```
-- LIST DEPARTMENT NAMES --
```

```
SELECT DISTINCT DEPARTMENT AS 'Departments'  
FROM HRdatacopy;
```

Departments
Admin Offices
Executive Office
IT/IS
Production
Sales
Software Engineering

```
-- NUMBER OF SPECIAL PROJECTS BY DEPARTMENT --
SELECT Department,
       SUM(SpecialProjectsCount) AS '# of Special Projects'
FROM HRdatacopy
GROUP BY Department;
```

Department	# of Special Projects
Admin Offices	33
Executive Office	0
IT/IS	296
Production	4
Sales	0
Software Engineering	46

```
-- NUMBER OF EMPLOYEES IN EACH DEPARTMENT (INCLUDING SEX DISTRIBUTION) --
```

```
SELECT Department,
       COUNT(EmpID) AS 'Number of Employees',
       COUNT(CASE
             WHEN Sex = 'F' THEN 1
             ELSE NULL
             END) AS 'Female',
       COUNT(CASE
             WHEN Sex = 'M' THEN 1
             ELSE NULL
             END) AS 'Male'
FROM HRdatacopy
GROUP BY Department
ORDER BY 'Number of Employees' DESC;
```

Department	Number of Employees	Female	Male
Production	209	126	83
IT/IS	50	22	28
Sales	31	15	16
Software Engineering	11	6	5
Admin Offices	9	6	3
Executive Office	1	1	0

-- DISTRIBUTION OF RACE DESCRIPTION IN EACH DEPARTMENT (ACTIVE EMPLOYEES) --

```

SELECT Department,
COUNT(EmpID) AS 'Number of Employees',
COUNT(CASE
  WHEN RaceDesc = 'American Indian or Alaska Native' THEN 1
  ELSE NULL
END) AS 'American Indian or Alaska Native',
COUNT(CASE
  WHEN RaceDesc = 'Asian' THEN 1
  ELSE NULL
END) AS 'Asian',
COUNT(CASE
  WHEN RaceDesc = 'Black or African American' THEN 1
  ELSE NULL
END) AS 'Black or African American',
COUNT(CASE
  WHEN RaceDesc = 'Hispanic' THEN 1
  ELSE NULL
END) AS 'Hispanic',
COUNT(CASE
  WHEN RaceDesc = 'Two or more races' THEN 1
  ELSE NULL
END) AS 'Two or more races',
COUNT(CASE
  WHEN RaceDesc = 'White' THEN 1
  ELSE NULL
END) AS 'White'
FROM HRdatacopy
WHERE EmploymentStatus = 'Active'
GROUP BY Department
ORDER BY 'Number of Employees' DESC;

```

Department	Number of Employees	American Indian or Alaska Native	Asian	Black or African American	Hispanic	Two or more races	White
Production	126	2	12	28	1	4	79
IT/IS	40	0	6	11	0	0	23
Sales	26	1	1	8	0	4	12
Software Engineering	7	0	1	1	0	0	5
Admin Offices	7	0	0	3	0	0	4
Executive Office	1	0	0	0	0	0	1

-- AVERAGE SALARIES IN EACH DEPARTMENT --

```
SELECT Department,
       CAST(AVG(Salary) AS INT) AS 'Avg. Salary'
FROM HRdatacopy
GROUP BY Department
ORDER BY 'Avg. Salary' DESC;
```

Department	Avg. Salary
Executive Office	250000
IT/IS	97064
Software Engineering	94989
Admin Offices	71791
Sales	69061
Production	59953

-- AVERAGE SALARIES IN EACH DEPARTMENT (ACCOUNTING FOR SEX) --

```
WITH m_salaries AS
    (SELECT Department,
             CAST(AVG(Salary) AS INT) AS 'Avg. Male Salary'
     FROM HRdatacopy
     WHERE Sex = 'M'
     GROUP BY Department),
     f_salaries AS
    (SELECT Department,
             CAST(AVG(Salary) AS INT) AS 'Avg. Female Salary'
     FROM HRdatacopy
     WHERE Sex = 'F'
     GROUP BY Department)
SELECT f_salaries.Department,
       ISNULL(m_salaries.[Avg. Male Salary],0) AS 'Avg. Male Salary',
       ISNULL(f_salaries.[Avg. Female Salary],0) AS 'Avg. Female Salary',
       ISNULL(([Avg. Male Salary] - [Avg. Female Salary]),0) AS 'Salary Difference'
FROM m_salaries
RIGHT OUTER JOIN f_salaries
ON m_salaries.Department = f_salaries.Department;
```

Department	Avg. Male Salary	Avg. Female Salary	Salary Difference
Admin Offices	73523	70926	2597
Executive Office	0	250000	0
IT/IS	99006	94593	4413
Production	60459	59620	839
Sales	66290	72016	-5726
Software Engineering	92689	96906	-4217



```
-- AVG. SALARIES OF POSITIONS EACH YEAR --
```

```
SELECT YEAR(DateofHire) AS 'Year',  
       Position,  
       ROUND(AVG(Salary),0) AS 'Avg. Salary'  
FROM HRdatacopy  
GROUP BY YEAR(DateofHire), Position  
ORDER BY 'Year';
```

```
-- DISPLAY CHANGES IN AVG. SALARIES FOR EACH POSITION OVER TIME --
```

```
SELECT *  
FROM  
    (SELECT Position,  
            YEAR(DateofHire) AS 'Year',  
            CAST(Salary AS INT) AS 'Salary'  
    FROM HRdatacopy) AS p  
PIVOT  
(  
    AVG(p.Salary)  
    FOR [Year] IN ([2006], [2007], [2008], [2009], [2010], [2011], [2012], [2013],  
[2014], [2015], [2016], [2017], [2018]))  
AS pivot_table
```

```
-- SALARY CONSIDERATIONS: DATASET DOES NOT ACCOUNT FOR EDUCATION (DEGREES, CERTS., ETC.)  
AND/OR TECHNICAL SKILLS (WHICH WOULD TRADITIONALLY FACTOR INTO SALARY) --  
/* BASIC RECRUITMENT INFORMATION */
```

```
-- LIST OF RECRUITMENT SOURCES --
```

```
SELECT DISTINCT RecruitmentSource  
FROM HRdatacopy;
```

RecruitmentSource
CareerBuilder
Diversity Job Fair
Employee Referral
Google Search
Indeed
LinkedIn
On-line Web application
Other
Website

-- NUMBER OF EMPLOYEES HIRED FROM 2006-2018 FROM EACH RECRUITMENT SOURCE WITH TOTALS --

```
SELECT YEAR(DateofHire) AS 'Year',
       COALESCE(RecruitmentSource, 'Total') AS 'Recruitment Source',
       COUNT(RecruitmentSource) AS 'Count'
FROM HRdatacopy
GROUP BY ROLLUP(YEAR(DateofHire), RecruitmentSource);
```

-- NUMBER OF EMPLOYEES RECEIVED FROM EACH RECRUITMENT SOURCE (ACCOUNTING FOR SEX DISTRIBUTION) --

```
SELECT RecruitmentSource,
       COUNT(EmpID) AS 'Employees Recruited',
       COUNT(CASE
            WHEN Sex = 'F' THEN 1 ELSE NULL
            END) AS 'Female',
       COUNT(CASE
            WHEN Sex = 'M' THEN 1 ELSE NULL
            END) AS 'Male'
FROM HRdatacopy
GROUP BY RecruitmentSource;
```

RecruitmentSource	Employees Recruited	Female	Male
CareerBuilder	23	16	7
Diversity Job Fair	29	15	14
Employee Referral	31	11	20
Google Search	49	33	16
Indeed	87	50	37
LinkedIn	76	44	32
On-line Web application	1	0	1
Other	2	1	1
Website	13	6	7

```
-- NUMBER OF EMPLOYEES RECEIVED FROM EACH RECRUITMENT SOURCE (ACCOUNTING FOR RACE
DESCRIPTION) --
```

```
SELECT RecruitmentSource,
COUNT(EmpID) AS 'Employees Recruited',
COUNT(CASE
WHEN RaceDesc = 'American Indian or Alaska Native' THEN 1
ELSE NULL
END) AS 'American Indian or Alaska Native',
COUNT(CASE
WHEN RaceDesc = 'Asian' THEN 1
ELSE NULL
END) AS 'Asian',
COUNT(CASE
WHEN RaceDesc = 'Black or African American' THEN 1
ELSE NULL
END) AS 'Black or African American',
COUNT(CASE
WHEN RaceDesc = 'Hispanic' THEN 1
ELSE NULL
END) AS 'Hispanic',
COUNT(CASE
WHEN RaceDesc = 'Two or more races' THEN 1
ELSE NULL
END) AS 'Two or more races',
COUNT(CASE
WHEN RaceDesc = 'White' THEN 1
ELSE NULL
END) AS 'White'
FROM HRdatacopy
GROUP BY RecruitmentSource;
```

RecruitmentSource	Employees Recruited	American Indian or Alaska Native	Asian	Black or African American	Hispanic	Two or more races	White
CareerBuilder	23	0	2	5	0	0	16
Diversity Job Fair	29	0	0	29	0	0	0
Employee Referral	31	0	1	5	0	0	25
Google Search	49	1	7	4	0	2	35
Indeed	87	1	10	17	1	4	54
LinkedIn	76	1	8	16	0	4	47
On-line Web application	1	0	0	0	0	0	1
Other	2	0	0	0	0	0	2
Website	13	0	1	4	0	1	7



Director of Sales	0	0	0	0	0	0	0	0	1	0	0	0	0
Enterprise Architect	0	0	0	0	0	0	0	0	1	0	0	0	0
IT Director	0	0	0	0	0	1	0	0	0	0	0	0	0
IT Manager - DB	0	0	0	0	0	0	1	1	0	0	0	0	0
IT Manager - Infra	0	0	0	0	0	0	1	0	0	0	0	0	0
IT Manager - Support	0	0	0	0	0	0	0	0	1	0	0	0	0
IT Support	0	0	0	0	1	2	1	0	0	4	0	0	0
Network Engineer	0	0	0	0	0	0	0	0	1	4	0	0	0
President & CEO	0	0	0	0	0	0	1	0	0	0	0	0	0
Principal Data Architect	0	0	0	0	0	0	0	0	0	1	0	0	0
Production Manager	0	0	0	1	2	4	2	2	1	1	1	0	0
Production Technician I	0	1	2	3	1	43	26	26	23	8	3	0	1
Production Technician II	0	1	0	1	3	22	5	9	11	2	3	0	0
Sales Manager	0	0	0	0	0	1	0	0	2	0	0	0	0
Senior BI Developer	0	0	0	0	0	0	0	0	0	0	1	2	0
Shared Services Manager	0	0	0	0	0	0	0	0	0	0	1	0	0
Software Engineer	0	0	0	0	0	3	2	3	2	0	0	0	0
Software Engineering Manager	0	0	0	0	0	1	0	0	0	0	0	0	0
Sr. Accountant	0	0	0	1	0	0	0	0	0	1	0	0	0
Sr. DBA	0	0	0	0	0	0	0	0	0	1	1	0	0
Sr. Network Engineer	0	0	0	0	0	0	0	0	3	1	1	0	0

```
-- NUMBER OF EMPLOYEES LOST FOR EACH POSITION EACH YEAR --
SELECT *
FROM
    (SELECT Position,
        YEAR(DateofTermination) AS 'Year',
        EmpID AS 'Number of Employees'
    FROM HRdatacopy) AS e
PIVOT
(
    COUNT(e.[Number of Employees])
    FOR [Year] IN ([2006], [2007], [2008], [2009], [2010], [2011], [2012], [2013],
    [2014], [2015], [2016], [2017], [2018]))
AS pivot_table
```

Position	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Accountant I	0	0	0	0	0	0	0	0	0	0	0	0	0
Administrative Assistant	0	0	0	0	0	0	0	1	0	1	0	0	0
Area Sales Manager	0	0	0	0	0	0	0	0	2	1	0	0	1
BI Developer	0	0	0	0	0	0	0	0	0	0	0	0	0
BI Director	0	0	0	0	0	0	0	0	0	0	0	0	0
CIO	0	0	0	0	0	0	0	0	0	0	0	0	0
Data Analyst	0	0	0	0	0	0	0	0	0	1	1	0	0
Data Architect	0	0	0	0	0	0	0	0	0	0	0	0	0
Database Administrator	0	0	0	0	0	0	0	0	0	2	0	1	0
Director of Operations	0	0	0	0	0	0	0	0	0	0	0	0	0
Director of Sales	0	0	0	0	0	0	0	0	0	0	0	0	0
Enterprise Architect	0	0	0	0	0	0	0	0	0	0	1	0	0
IT Director	0	0	0	0	0	0	0	0	0	0	0	0	0
IT Manager - DB	0	0	0	0	0	0	0	0	0	1	0	0	0
IT Manager - Infra	0	0	0	0	0	0	0	0	0	0	0	0	0
IT Manager - Support	0	0	0	0	0	0	0	0	0	0	0	0	0
IT Support	0	0	0	0	0	0	0	0	0	0	0	0	0
Network Engineer	0	0	0	0	0	0	0	0	0	0	1	0	0
President & CEO	0	0	0	0	0	0	0	0	0	0	0	0	0
Principal Data Architect	0	0	0	0	0	0	0	0	0	0	0	0	1
Production Manager	0	0	0	0	0	0	2	0	1	1	1	0	0
Production Technician I	0	0	0	0	0	0	2	3	5	11	15	6	10

Production Technician II	0	0	0	0	1	3	4	8	4	3	2	1	0
Sales Manager	0	0	0	0	0	0	0	0	1	0	0	0	0
Senior BI Developer	0	0	0	0	0	0	0	0	0	0	0	0	0
Shared Services Manager	0	0	0	0	0	0	0	0	0	0	0	0	0
Software Engineer	0	0	0	0	0	0	0	1	0	2	0	0	1
Software Engineering Manager	0	0	0	0	0	0	0	0	0	0	0	0	0
Sr. Accountant	0	0	0	0	0	0	0	0	0	0	0	0	0
Sr. DBA	0	0	0	0	0	0	0	0	0	0	1	0	0
Sr. Network Engineer	0	0	0	0	0	0	0	0	0	0	0	0	0

-- CALCULATING COMPANY TENURES FOR INACTIVE EMPLOYEES (ONLY THOSE WHO LEFT THE COMPANY FOR RETENTION PURPOSES)--

-- ACCOUNTING FOR RACE DESCRIPTION DISTRIBUTION --

```

SELECT DATEDIFF(YEAR, DateofHire, DateofTermination) AS 'Years of Service',
       COUNT(DATEDIFF(YEAR, DateofHire, DateofTermination)) AS 'Number of Employees',
       COUNT(CASE
         WHEN RaceDesc = 'American Indian or Alaska Native' THEN 1
         ELSE NULL
         END) AS 'American Indian or Alaska Native',
       COUNT(CASE
         WHEN RaceDesc = 'Asian' THEN 1
         ELSE NULL
         END) AS 'Asian',
       COUNT(CASE
         WHEN RaceDesc = 'Black or African American' THEN 1
         ELSE NULL
         END) AS 'Black or African American',
       COUNT(CASE
         WHEN RaceDesc = 'Hispanic' THEN 1
         ELSE NULL
         END) AS 'Hispanic',
       COUNT(CASE
         WHEN RaceDesc = 'Two or more races' THEN 1
         ELSE NULL
         END) AS 'Two or more races',
       COUNT(CASE
         WHEN RaceDesc = 'White' THEN 1
         ELSE NULL
         END) AS 'White'
FROM HRdatacopy
WHERE EmploymentStatus = 'Voluntarily Terminated'
GROUP BY DATEDIFF(YEAR, DateofHire, DateofTermination);

```

Years of Service	Number of Employees	American Indian or	Asian	Black or African American	Hispanic	Two or more races	White

		Alaska Native					
0	2	0	0	2	0	0	0
1	18	0	5	2	0	1	10
2	10	0	0	1	0	1	8
3	20	0	0	5	0	0	15
4	13	0	1	7	0	0	5
5	11	0	0	4	0	0	7
6	6	0	0	2	0	0	4
7	7	0	3	0	0	0	4
9	1	0	0	0	0	0	1

-- CALCULATING COMPANY TENURES FOR INACTIVE EMPLOYEES (ONLY THOSE WHO LEFT THE COMPANY FOR RETENTION PURPOSES)--

-- ACCOUNTING FOR SEX DISTRIBUTION --

```

SELECT DATEDIFF(YEAR, DateofHire, DateofTermination) AS 'Years of Service',
       COUNT(DATEDIFF(YEAR, DateofHire, DateofTermination)) AS 'Number of Employees',
       COUNT(CASE
         WHEN Sex = 'F' THEN 1
         ELSE NULL
         END) AS 'Female',
       COUNT(CASE
         WHEN Sex = 'M' THEN 1
         ELSE NULL
         END) AS 'Male'
FROM HRdatacopy
WHERE EmploymentStatus = 'Voluntarily Terminated'
GROUP BY DATEDIFF(YEAR, DateofHire, DateofTermination);

```

Years of Service	Number of Employees	Female	Male
0	2	1	1
1	18	9	9
2	10	7	3
3	20	12	8
4	13	8	5
5	11	5	6
6	6	5	1
7	7	4	3
9	1	0	1



```
-- DEPARTURE REASONS FOR EMPLOYEES THAT VOLUNTARILY CHOSE TO LEAVE --
```

```
SELECT TermReason,
       COUNT(TermReason) AS 'Count'
FROM HRdatacopy
WHERE EmploymentStatus = 'Voluntarily Terminated'
GROUP BY TermReason
ORDER BY 'Count' DESC;
```

TermReason	Count
Another position	20
unhappy	14
more money	11
career change	9
hours	8
relocation out of area	5
return to school	5
military	4
retiring	4
maternity leave - did not return	3
medical issues	3
performance	1
attendance	1

```
-- DEPARTURE REASONS FOR EMPLOYEES THAT VOLUNTARILY CHOSE TO LEAVE (SEGMENTED BY SEX) --
```

```
WITH women_TermReason AS(
    SELECT TermReason,
           COUNT(TermReason) AS 'Count'
    FROM HRdatacopy
    WHERE EmploymentStatus = 'Voluntarily Terminated' AND Sex = 'F'
    GROUP BY TermReason),
    men_TermReason AS(
    SELECT TermReason,
           COUNT(TermReason) AS 'Count'
    FROM HRdatacopy
    WHERE EmploymentStatus = 'Voluntarily Terminated' AND Sex = 'M'
    GROUP BY TermReason)
SELECT m.TermReason AS 'Reasons for Voluntary Departure',
       ISNULL(w.[Count],0) AS 'Women',
       ISNULL(m.[Count],0) AS 'Men',
       (ISNULL(w.[Count],0) + ISNULL(m.[Count],0)) AS 'Total'
FROM women_TermReason AS w
RIGHT OUTER JOIN men_TermReason AS m
ON w.TermReason = m.TermReason
GROUP BY m.TermReason, w.[Count], m.[Count]
ORDER BY m.TermReason;
```

Reasons for Voluntary Departure	Women	Men	Total
Another position	16	4	20
attendance	0	1	1
career change	4	5	9
hours	3	5	8
maternity leave - did not return	2	1	3
medical issues	1	2	3
military	2	2	4
more money	6	5	11
performance	0	1	1
relocation out of area	3	2	5
retiring	1	3	4
return to school	4	1	5
unhappy	9	5	14

-- FINDING THE DEPARTMENTS WHERE THE TOP 3 REASONS FOR LEAVING OCCUR --

```

SELECT Department,
       TermReason,
       COUNT(TermReason) AS 'Count'
FROM HRdatacopy
WHERE TermReason IN ('Another position', 'unhappy', 'more money')
GROUP BY Department, TermReason
ORDER BY 'Count' DESC;

```

Department	TermReason	Count
Production	Another position	17
Production	unhappy	14
Production	more money	11
IT/IS	Another position	1
Sales	Another position	1
Software Engineering	Another position	1

-- AVERAGE ENGAGEMENT, SATISFACTION AND PERFORMANCE SCORES (SEGMENTED BY SEX) --

```
SELECT Sex,
       ROUND(AVG(EngagementSurvey),2) AS 'avg_engagement',
       ROUND(AVG(EmpSatisfaction), 2) AS 'avg_satisfaction',
       ROUND(AVG(PerfScoreID),2) AS 'avg_performance'
FROM HRdatacopy
GROUP BY Sex;
```

Sex	avg_engagement	avg_satisfaction	avg_performance
F	4.14	3.93	3.01
M	4.08	3.84	2.94

-- AVERAGE ENGAGEMENT, SATISFACTION AND PERFORMANCE SCORES BY DEPARTMENT --

```
SELECT Department,
       ROUND(AVG(EngagementSurvey),2) AS 'avg_engagement',
       ROUND(AVG(EmpSatisfaction),2) AS 'avg_satisfaction',
       ROUND(AVG(PerfScoreID),2) AS 'avg_performance'
FROM HRdatacopy
GROUP BY Department
```

Department	avg_engagement	avg_satisfaction	avg_performance
Admin Offices	4.39	3.56	3
Executive Office	4.83	3	3
IT/IS	4.15	3.96	3.06
Production	4.13	3.86	2.97
Sales	3.82	4.03	2.84
Software Engineering	4.06	4.09	3.09

### Analysis Points:

- The top 3 recruitment sources were Indeed (87), LinkedIn (76) and Google Search (49). Note: The recruitment sources Google Search, Website, and On-line Web application are ambiguous as these could be grouped into the other sources in one form or another. Recommendation to HR would be to evaluate whether these sources can be consolidated.
- The only employees sourced from Diversity Job Fair were Black or African Americans (29, Women-15, Men-14). Recommendation for HR Recruitment: Reevaluate the planning and execution of these events to determine why they do not attract potential hires from other diverse/ethnic groups (particularly those from groups that had low counts across all other recruitment sources (American Indian or Alaska Native and Hispanic).
- The top 3 reasons for voluntary employee departure were '*Another position*' (20, Women-16, Men-4), 'unhappy' (14, Women-9, Men-5) and 'more money' (11, Women-6, Men-5). Note: Term Reason considerations: 'Another position' and 'career change' could be considered the same, as 'Another position' could indicate moving to a different position within the company (e.g., promotion, lateral). Not enough information in the dataset to draw conclusions.
- Of the employees that voluntarily left the company, the highest numbers departed at the 3-year mark (20, Women-12, Men-8), next at the 1-year mark (18, Women-9, Men-9) and finally the 4-year mark (13, Women-8, Men-5). Recommendations for Retention: these year-marks could be used in conjunction with the reasons for voluntary departure to get employees to stay before they consider leaving. Ex. Prior to when an employee reaches the 3-year and 4-year marks (perhaps 6 months before), open a dialogue with them on whether they are interested in lateral or promotional opportunities.

## CORRELATION:

### *Hypotheses –*

1. Hypothesis 1 – There is a high correlation between Employee Satisfaction and Engagement.
2. Hypothesis 2 – There is a high correlation between Employee Satisfaction and Performance.
3. Hypothesis 3 – There is a high correlation between Engagement and Performance.

-----

/\* CORRELATION \*/

/\* 1ST CORRELATION SET - ATTEMPT TO DETERMINE IF THERE IS A CORRELATION BETWEEN EMPLOYEE SATISFACTION AND ENGAGEMENT SURVEY \*/

-- CORRELATION BETWEEN EMPLOYEE SATISFACTION AND ENGAGEMENT SURVEY SCORE --

```
SELECT ((AVG(EngagementSurvey * EmpSatisfaction)) - (AVG(EngagementSurvey) *  
AVG(EmpSatisfaction)))/  
((STDEVP(EngagementSurvey) * (STDEVP(EmpSatisfaction)))) AS 'pearson_coefficient'  
FROM HRdatacopy;
```

pearson_coefficient
0.187105

/\* 2ND CORRELATION SET - ATTEMPT TO DETERMINE IF THERE IS A CORRELATION BETWEEN EMPLOYEE SATISFACTION AND PERFORMANCE SCORE \*/

-- CORRELATION BETWEEN PERFORMANCE SCORE AND EMPLOYEE SATISFACTION --

```
SELECT ((AVG(PerfScoreID * EmpSatisfaction)) - (AVG(PerfScoreID) *  
AVG(EmpSatisfaction)))/  
((STDEVP(PerfScoreID) * (STDEVP(EmpSatisfaction)))) AS 'pearson_coefficient'  
FROM HRdatacopy;
```

pearson_coefficient
0.303579382

/\* 3RD CORRELATION SET - ATTEMPT TO DETERMINE IF THERE IS A CORRELATION BETWEEN PERFORMANCE SCORE AND ENGAGEMENT SURVEY \*/

-- CORRELATION BETWEEN PERFORMANCE SCORE AND ENGAGEMENT SURVEY SCORE --

```
SELECT ((AVG(PerfScoreID * EngagementSurvey)) - (AVG(PerfScoreID) *  
AVG(EngagementSurvey)))/  
((STDEVP(PerfScoreID) * (STDEVP(EngagementSurvey)))) AS 'pearson_coefficient'  
FROM HRdatacopy;
```

pearson_coefficient
0.544926678

**Correlation Points:**

Hypothesis 1 – There is a high correlation between Employee Satisfaction and Engagement. **Result** – The correlation coefficient was 0.187105. This shows a negligible relationship between Employee Satisfaction and Engagement.

**Hypothesis 2** – There is a high correlation between Employee Satisfaction and Performance. **Result** – The correlation coefficient was 0.303579382. This shows a weak relationship between Employee Satisfaction and Performance.

**Hypothesis 3** – There is a high correlation between Engagement and Performance. **Result** – The correlation coefficient was 0.544926678. This shows a moderate relationship between Engagement and Performance.

*Note on dataset regarding salary: Dataset does not account for education (degrees, certs., etc.) and/or technical skills (which would traditionally factor into salary).*