SCHOOL OF ENGINEERING AND TECHNOLOGY

ASSIGNMENT SUBMISSION FORM

PROGRAMME : BSBA / BSDA
SEMESTER : September 2024
SUBJECT : IST2034 Analytics Engineering
DEADLINE : 27 December 2024 11:59 PM MYT

INSTRUCTIONS TO CANDIDATES

- This assignment will contribute 30% to your final grade.
- This is a group assignment. Refer to the Assignment Brief for more details.

IMPORTANT

The University requires students to adhere to submission deadlines for any form of assessment. Penalties are applied in relation to unauthorized late submission of work.

- Coursework submitted after the deadline but within 1 week will be accepted for a maximum mark of 40%.
- Work handed in following the extension of 1 week after the original deadline will be regarded as a non-submission and marked zero.

Student's Declaration:

	No.	Student Name	Student ID	Signature	Date
We,	**1.**	Ayu Wen Li	22017867	Wenli	26/12/2024
	2.	Goo Weng Xi	22034987	*wengxi*	26/12/2024
	3.	Keertana A/P Subramaniam	23109614	*keertana*	26/12/2024
	4.	Lim Phin Han	22052898	*phinhan*	26/12/2024
	5.	Siow Qi Yung	22053037	*qiyung*	26/12/2024
	6.	Tham Yan Qi	22055719	tham	26/12/2024

received the assignment and read the comments.

Academic Honesty Acknowledgement

"We ***(as per the Student Names stated above)*** verify that this paper contains entirely our own work. We have not consulted with any outside person or materials other than what was specified (an interviewee, for example) in the assignment or the syllabus requirements. Further, we have not copied or inadvertently copied ideas, sentences, or paragraphs from another student. We realise the penalties (refer to page 16, 5.5, Appendix 2, page 44 of the student handbook diploma and undergraduate programme) for any kind of copying or collaboration on any assignment."

1. Ayu Wen Li
2. Goo Weng Xi
3. Keertana A/P Subramaniam
4. Lim Phin Han
5. Siow Qi Yung
6. Tham Yan Qi

Wenli *wengxi keertana phinhan qiyung* tham 26/12/2024
…..... (Student's signature / Date)

Data Protection

The protection of personal data is an important concern to Sunway University and any personal data collected on this form will be treated in accordance with the Personal Data Protection Notice of the institution.
http://sunway.edu.my/pdpa/notice_english (English version)
http://sunway.edu.my/pdpa/notice_bm (Malay version)

Group Acknowledgement

We, the undersigned members of Group 13, hereby acknowledge and agree to the following regarding our individual contributions to the group assignment.

Group Information:

- Group Number: 13
- Group Members:
  1. AYU WEN LI - 22017867
  2. GOO WENG XI - 22034987
  3. KEERTANA A/P SUBRAMANIAM - 23109614
  4. LIM PHIN HAN - 22052898
  5. SIOW QI YUNG - 22053037
  6. THAM YAN QI - 22055719

Equal Participation:
We confirm that each member of the group has actively and equally participated in all aspects of the assignment, including data collection, analysis, and report writing.

Recognising Unique Contributions:
We acknowledge and appreciate the unique strengths, skills, and perspectives that each member brought to the group, contributing to the overall success of the assignment.

Fair Distribution of Work:
We affirm that the workload and responsibilities were distributed equitably among all group members, allowing each member to make a meaningful contribution.

Verification of Individual Contributions:
Each member verifies and agrees that the content submitted under their name is a result of their direct involvement and effort in the assignment. We understand that the submitted work reflects our individual and collective efforts and represents the collaborative outcome of the entire group.

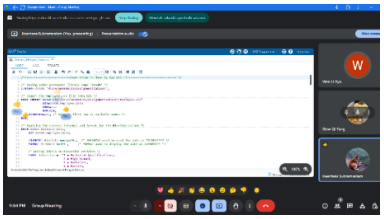
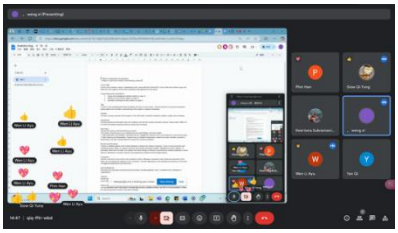
Commitment to Collaboration:
We are committed to open communication and collaboration within the group.

Dispute Resolution:
In the event of any disputes or concerns, the group agrees to address and resolve the issues through open communication and collaboration in a constructive manner.

Signatures:
By signing below, each group member verifies and acknowledges the individual contributions made by themselves and their fellow group members in the completion of the assignment.

Student 1:	Wenli	Date:	26/12/2024
Student 2:	wengxi	Date:	26/12/2024
Student 3:	keertana	Date:	26/12/2024
Student 4:	phinhan	Date:	26/12/2024
Student 5:	qiyung	Date:	26/12/2024
Student 6:	tham	Date:	26/12/2024

Meeting Records

Date	Attended by	Items Discussed
5/12 2-4pm  	ALL	- Decide dataset searched by every group member - Deciding on research questions - separating tasks and research questions - setting internal deadlines
9/12 8:00 – 9:00 pm 	Ayu Wen Li, Siow Qi Yung, Keertana A/P Subramaniam	- Doing Data Pre-Processing and Validation, imported the raw datasets, merged it, and checked for errors.
10/12 2:30- 3:30pm 	Goo Weng Xi, Lim Phin Han, Tham Yan Qi	- Doing Data Cleaning and Manipulation handled missing data, created secondary variables, grouped or binned data, and other transformations to answer the research questions.
12/12 2-4pm 	ALL	- doing research questions (coding) and research & discussion - checking for progress - troubleshooting and addressing any problems encountered
19/12 11am-1230pm 	ALL	- writing abstract and conclusion - deciding on the main title - organising reference lists - appendix - finalising

Individual Contribution

Section	Worked on by
Introduction	Ayu Wen Li, Siow Qi Yung
Abstract	Lim Phin Han, Keertana A/P Subramaniam
Data Exploration and Validation	Ayu Wen Li, Siow Qi Yung, Keertana A/P Subramaniam
Data Cleaning and Manipulation	Goo Weng Xi, Lim Phin Han, Tham Yan Qi
Research Question Q1 (code)	Ayu Wen Li (main), Siow Qi Yung (assisting)
Research Question Q2 (code)	Keertana A/P Subramaniam (main), Siow Qi Yung (assisting)
Research Question Q3 (code)	Lim Phin Han (main), Goo Weng Xi (assisting)
Research Question Q4 (code)	Tham Yan Qi (main), Goo Weng Xi (assisting)
Result and Analysis Discussion for Q1	Ayu Wen Li, Siow Qi Yung
Result and Analysis Discussion for Q2	Keertana A/P Subramaniam, Siow Qi Yung
Result and Analysis Discussion for Q3	Lim Phin Han, Goo Weng Xi
Result and Analysis Discussion for Q4	Tham Yan Qi, Goo Weng Xi
Conclusion	Goo Weng Xi, Tham Yan Qi

"Analyzing Workplace Dynamics: Factors Influencing Employees' Performance, Satisfaction, and Training Opportunities Across Departments and Genders"

Ayu Wen Li
22017867

Goo Weng Xi
22034987

Keertana A/P Subramaniam
23109614

Lim Phin Han
22052898

Siow Qi Yung
22053037

Tham Yan Qi
22055719

Abstract This study aims to investigate the key factors influencing employee performance, satisfaction, and development within Atlas Lab, examining variations across different departments and genders. This research has adopted data from Kaggle's "HR Analytics: Employee Attrition & Performance" dataset, focusing on the 'Employee.csv' and 'PerformanceRating.csv' files, which included information on 1,470 employees. The research addresses four primary questions: (1) Do manager ratings differ across departments? (2) Is there gender-based disparities in performance ratings? (3) Which factor—work-life balance, relationship satisfaction, environment satisfaction, or salary—most strongly impacts job satisfaction? and (4) How does access to training opportunities vary among departments? This study uses descriptive analysis to summarise key variables, ANOVA to test differences in manager ratings across departments, and t-tests to compare performance ratings between genders. Correlation analysis is also used to explore the relationships between job satisfaction and factors such as work-life balance, relationship satisfaction, environmental satisfaction, and salary. Moreover, this study revealed that manager ratings and access to training opportunities are consistent across departments, reflecting equitable resource distribution. Additionally, no significant differences were found in performance ratings between male and female employees, highlighting fairness in evaluations. Besides, environmental satisfaction has a weak but significant relationship with job satisfaction in 2021, whereas no significant relationships were observed in 2016. In conclusion, Atlas Lab ensured fairness in its performance evaluation and resource allocation. However, this study also has some limitations, such as relying on only two years of data and the weak correlations observed, which suggests that there may be other unmeasured factors influencing the result. Future research should consider using longitudinal data and exploring other variables to gain a deeper understanding of employee satisfaction. In addition, SAS programming is essential in managing data and conducting statistical analysis, proving to be an effective tool for gaining valuable insights from HR datasets.

INTRODUCTION

The dataset is from Kaggle which is titled "HR Analytics: Employee Attrition & Performance". 'Employee.csv' dataset and 'PerformanceRating.csv' dataset was used in this study for further exploration and research. There is a total of 5568 data after merging and cleaning these two datasets. The Employee dataset contains detailed data of employees that work in Atlas Lab. The variables in this dataset are:

- EmployeeID: Unique identifier of each employee.
- FirstName: The first name of the employee.
- LastName: The last name of the employee.
- Gender: The gender of the employee.
- Age: The age of the employee.
- BusinessTravel: The frequency of business travel for the employee.
- Department: The department in which the employee works.
- DistanceFromHome (KM): The distance between the employee's home and workplace in kilometres.
- State: The state in which the employee resides.
- Ethnicity: The ethnicity of the employee.
- MaritalStatus: The marital status of the employee.
- Salary: The annual salary of the employee.
- StockOptionLevel: The level of stock options granted to the employee.
- OverTime: Whether the employee works overtime (Yes/No).
- HireDate: The date the employee was hired.
- Attrition: Whether the employee has left the company (Yes/No).
- YearsAtCompany: The number of years the employee has been with the company.
- YearsInMostRecentRole: The number of years the employee has been in their most recent role.
- YearsSinceLastPromotion: The number of years since the employee's last promotion.
- YearsWithCurrManager: The number of years the employee has worked with their current manager.

PerformanceRating dataset contains the data about the performance of the employees in the company. The variables in this dataset are:

- PerformanceID: Unique identifier for each performance review.
- EmployeeID: Unique identifier for the employee being reviewed.
- ReviewDate: The date of the performance review.
- EnvironmentSatisfaction: Rating of the employee's satisfaction with their work environment.
- JobSatisfaction: Rating of the employee's satisfaction with their job.
- RelationshipSatisfaction: Rating of the employee's satisfaction with workplace relationships.
- TrainingOpportunitiesWithinYear: Number of training opportunities available to the employee within the year.
- TrainingOpportunitiesTaken: Number of training opportunities the employee has taken.
- WorkLifeBalance: Rating of the employee's work-life balance.
- SelfRating: The employee's self-assessment rating.
- ManagerRating: The manager's rating of the employee's performance.

Four research questions have been created by using the merging of the two datasets. Firstly, we aim to analyse how manager ratings of employees vary across different departments. Our objective is to compare the distribution of performance ratings across different departments to identify which department its employees perform better, worse, or whether every department has similar performance ratings. If any department has very low ratings, the company can discuss strategies to make improvements for the department.

Secondly, we aim to examine if there are any significant differences in performance ratings between male and female employees. Our purpose is to analyse whether there is gender-based disparities in employee performance ratings and to understand whether gender has an impact on performance ratings. The employee salary and performance ratings including work environment satisfaction, job satisfaction, workplace relationship satisfaction and work-life balance are the key variables for conducting tests in this research.

Other than that, we aimed to investigate if work-life balance, relationship satisfaction, environment satisfaction or salary has the strongest influence on job satisfaction of employees. Our objective is to identify the most influential factors affecting employees' satisfaction with their job to help the company to identify and focus on the most impactful factor to improve the job satisfaction of employees and their working motivation.

Lastly, analysing the variation of training opportunities across each department could help management to identify the departments with less and limited training resources. Therefore, the company can focus on providing more training opportunities for the employees to enhance their performance.

DATA ANALYSIS AND INSIGHTS

A. Research Questions

(1) How do manager ratings vary across different departments?
(2) Are there significant differences in performance ratings between male and female employees?
(3) Which factor—work-life balance, relationship satisfaction, environment satisfaction or salary—has the strongest impact on job satisfaction?
(4) How does access to training opportunities vary across departments?

B. Data Pre-Processing and Validation:

Firstly, the raw datasets named Employee.csv and PerformanceRating.csv was imported using proc import statement before proceeding with the data pre-processing and validation [1]. The data pre-processing was started by assigning correct labels for the variables from both Employee and PerformanceRating datasets such as education, environment satisfaction, job satisfaction, relationship satisfaction, work-life balance, self-rating and manager rating for clarity based on the label given by the author. This is because the author of the dataset has replaced the ordinal variables mentioned above with numbers from 1-5.

Then, both data were formatted and merged by their common variable which is EmployeeID to ensure that each performance rating records are aligned with the employee's information. Moving on, observations with missing values for performance rating were identified and removed to maintain data accuracy and consistency when performing data analysis in the future. However, those removed observations were not deleted but were moved to another dataset to ensure that employees' information would still be available and could be found in the company's data in future when needed. It is observed that employees who do not submit any performance rating (missing values) have YearsAtCompany less than or equal to one. This may be due to the reasons that the duration of working is too short that it has not reached the date for submitting a performance rating or they might still be in probation period.

Following, duplicated data were found if data was sorted by solely EmployeeID. However, after data understanding and exploration, it is observed that duplicate records were caused by the same employees submitting the performance ratings annually. Therefore, it was identified and listed in another dataset for clarity but not removed from the merged dataset as they are necessary for conducting employee performance rating analysis in subsequent steps.

Data cleaning and validation are done by performing data integrity checks. The hire and performance review dates were checked to verify that the performance review dates always come after the corresponding hire dates of the employee. Any records that failed to meet this criterion were excluded from the analysis due to potential data collection errors.

Last but not least, simple summary statistics were generated for both numeric and categorical variables to detect potential anomalies and outliers. For example, it was checked that there are no negative values for variables like Age and YearsAtCompany. Also, it is confirmed that the company is not hiring under-aged employees as the minimum age of employees in the dataset is 18.

Throughout each data processing step, the resulting dataset was exported into **sas7bdat** format and saved in a permanent library **ASSGN** to ensure efficient use in future analysis. Finally, the cleaned sorted and merged data will be used in future analysis in answering subsequent research questions.

RESULTS AND DISCUSSION

(1) How do manager ratings vary across different departments?

Descriptive Statistics for Manager Ratings by Department

The MEANS Procedure

Analysis Variable : ManagerRating 1 = Unacceptable, 2 = Needs Improvement, 3 = Meets Expectation, 4 = Exceeds Expectation, 5 = Above and Beyond

Department	N Obs	Mean	Std Dev	Minimum	Maximum
Human Resources	248	3.4354839	1.0120083	2.0000000	5.0000000
Sales	1728	3.4392361	0.9725889	2.0000000	5.0000000
Technology	3592	3.4877506	0.9486189	2.0000000	5.0000000

Fig. 1: Results showing Descriptive Statistics for Manager Ratings by Department

To analyze and compare the distribution of Manager Ratings across various departments within the organization, descriptive statistics was generated using the PROC MEANS statement to measure the means, standard deviation, minimum and maximum for the average manager ratings of each department. The results show that the Technology Department has the highest average ratings (3.488) across all departments, second is the Sales Department (3.439) and the Human Resources Department has the lowest average ratings (3.435). This indicates that the average employees' performance ratings by manager are relatively similar across the three different departments with minor differences.

Similarly, the standard deviation also shows little variability across the departments. The Human Resources Department shows the highest standard deviation of 1.012, while the Sales Department has a standard deviation of 0.973 and the Technology Department shows the smallest standard deviation of 0.948. The results suggest that the manager's rating in the Human Resources Department has the highest variability while the Technology Department has the least spread-out manager rating. However, it is also important to note that the number of observations across the three departments shows significant difference, with the results showing 248 observations (Human Resources), 1728 (Sales) and 3592 (Technology) due to the difference in number of employees. It may also be due to a greater number of employees from the technology and sales department having longer employment periods, therefore receiving multiple performance ratings from their manager. The difference in observations is an important aspect as it can affect the results significantly.


ANOVA for Employee Performance Rating by Manager Across Departments

The ANOVA Procedure

Class Level Information

Class	Levels	Values
Department	3	Human Resources Sales Technology

Number of Observations Read	5568
Number of Observations Used	5568

ANOVA for Employee Performance Rating by Manager Across Departments

The ANOVA Procedure

Dependent Variable: ManagerRating 1 = Unacceptable, 2 = Needs Improvement, 3 = Meets Expectation, 4 = Exceeds Expectation, 5 = Above and Beyond

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3.061894	1.530947	1.66	0.1894
Error	5565	5118.048558	0.919685		
Corrected Total	5567	5121.110453			

R-Square	Coeff Var	Root MSE	ManagerRating Mean
0.000598	27.63403	0.959002	3.470366

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Department	2	3.06189449	1.53094724	1.66	0.1894

Fig. 2: Results showing ANOVA for Employee Performance Rating by Manager Across Departments.

Additionally, the research question has been tested using the Analysis of Variance test (ANOVA) to test whether performance ratings vary across different departments. It is assumed that the distribution is normal, and the significance level is at 5% before performing the ANOVA test.

The results show that the p-value is 0.1894, which is greater than 0.05 (5% significance level). Therefore, there is no statistically significant difference in ratings from managers across the departments (Human Resources, Sales, and Technology). Since the average ratings are approximately above 3, this may indicate that every employee across the different departments is performing relatively well and has met the expectations of managers. It may suggest that the company's workplace environment is fair, and it provides comparable opportunities, training resources, and workload for each department, allowing their employees to perform moderately well. Moreover, this may also suggest

that the manager's standard and expectation for employee's performance are comparatively similar across the three departments.

ANOVA for Employee Performance Rating by Manager Across Departments

The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for ManagerRating

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	5565
Error Mean Square	0.919685
Critical Value of Studentized Range	3.31538

Comparisons significant at the 0.05 level are indicated by ***.

Department Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
Technology - Sales	0.04851	-0.01730	0.11433
Technology - Human Resources	0.05227	-0.09534	0.19987
Sales - Technology	-0.04851	-0.11433	0.01730
Sales - Human Resources	0.00375	-0.14891	0.15642
Human Resources - Technology	-0.05227	-0.19987	0.09534
Human Resources - Sales	-0.00375	-0.15642	0.14891

Fig. 3: Results showing Tukey's Studentized Range (HSD) Test

In addition to that, Tukey's Studentized Range (HSD) Test was also used to determine exactly which departments' performance ratings are different through pairwise comparisons. Based on the results, it is further confirmed that there are no statistically significant differences between the performance ratings for any pair of departments.

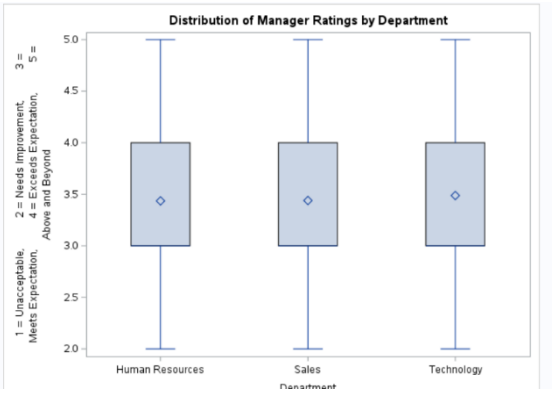


Fig. 4: Results showing Box Plot titled" Distribution of Manager Ratings by Department".

A visualization of the distribution of manager rating data was shown. The box plot illustrates that manager ratings across the departments are normally distributed with similar range, spread, median and mean. with no outliers shown. In conclusion, these results are aligned with the descriptive statistics and ANOVA results that there is no significant difference in manager rating across the departments The similar rating results could be an indication of even distribution of resources such as training opportunities and programs, ensuring fair opportunity for improved job performance and progress. It also shows that the managers across the departments have similar and standardized performance evaluation criteria, leading to consistent and fair assessments.

(2) Are there any significant differences in performance ratings (employee's salary, work environment satisfaction, job satisfaction, workplace relationship satisfaction, and work-life balance) between male and female employees?

Descriptive Statistics of Satisfaction Ratings and Salary by Gender

The MEANS Procedure

Gender	N Obs	Variable	N	Mean	Std Dev
Female	2603	EnvironmentSatisfaction	2603	3.8524779	0.9424956
		JobSatisfaction	2603	3.4429504	1.1762899
		RelationshipSatisfaction	2603	3.4391087	1.1583170
		WorkLifeBalance	2603	3.4068383	1.1316423
		Salary	2603	113107.48	96829.23
Male	2417	EnvironmentSatisfaction	2417	3.8622259	0.9579575
		JobSatisfaction	2417	3.4087712	1.1415059
		RelationshipSatisfaction	2417	3.3880844	1.1662289
		WorkLifeBalance	2417	3.4087712	1.1680299
		Salary	2417	116399.00	98117.79

Fig. 4: Results showing Descriptive Statistics of Satisfaction Ratings and Salary by Gender

We generated descriptive statistics using the PROC MEANS statement to measure the means, standard deviation, minimum and maximum. The result shows that the average salary of male employees (116399) is slightly higher than the average salary of female employees (113107). The male employees have higher average ratings of satisfaction with the working environment (3.86) than the female employees (3.85). However, the female employees have higher average ratings of satisfaction with their job (3.44 > 3.41) and workplace relationships (3.44 >3.39) than the male employees. Both female and male employees have almost the same average rating for their satisfaction with work-life balance (3.41).

Variable: Salary

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		2603	113107	96829.2	1897.9	22089.0	546549
Male		2417	116399	98117.8	1995.8	20387.0	547204
Diff (1-2)	Pooled		-3291.5	97451.8	2752.7		
Diff (1-2)	Satterthwaite		-3291.5		2754.1		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		113107	109386	116829	96829.2	94268.7	99533.8
Male		116399	112485	120313	98117.8	95427.8	100965
Diff (1-2)	Pooled	-3291.5	-8688.1	2105.1	97451.8	95582.1	99396.6
Diff (1-2)	Satterthwaite	-3291.5	-8690.8	2107.7			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	5018	-1.20	0.2319
Satterthwaite	Unequal	4980	-1.20	0.2321

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2416	2602	1.03	0.5078

Fig. 5: Results showing Pooled t-test for Salary

T-Test for Satisfaction Ratings and Salary Between Male and Female Employees

The TTEST Procedure

Variable: EnvironmentSatisfaction (1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied)

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		2603	3.8525	0.9425	0.0185	1.0000	5.0000
Male		2417	3.8622	0.9580	0.0195	1.0000	5.0000
Diff (1-2)	Pooled		-0.00975	0.9500	0.0268		
Diff (1-2)	Satterthwaite		-0.00975		0.0269		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		3.8525	3.8163	3.8887	0.9425	0.9176	0.9688
Male		3.8622	3.8240	3.9004	0.9580	0.9317	0.9858
Diff (1-2)	Pooled	-0.00975	-0.0624	0.0429	0.9500	0.9317	0.9689
Diff (1-2)	Satterthwaite	-0.00975	-0.0624	0.0429			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	5018	-0.36	0.7164
Satterthwaite	Unequal	4977.3	-0.36	0.7166

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2416	2602	1.03	0.4150

Fig. 6: Results showing Pooled t-test for EnvironmentSatisfaction

Variable: JobSatisfaction (1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied)

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		2603	3.4430	1.1763	0.0231	1.0000	5.0000
Male		2417	3.4088	1.1415	0.0232	1.0000	5.0000
Diff (1-2)	Pooled		0.0342	1.1597	0.0328		
Diff (1-2)	Satterthwaite		0.0342		0.0327		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		3.4430	3.3977	3.4882	1.1763	1.1452	1.2091
Male		3.4088	3.3632	3.4543	1.1415	1.1102	1.1746
Diff (1-2)	Pooled	0.0342	-0.0300	0.0984	1.1597	1.1374	1.1828
Diff (1-2)	Satterthwaite	0.0342	-0.0300	0.0983			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	5018	1.04	0.2968
Satterthwaite	Unequal	5008.2	1.04	0.2963

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2602	2416	1.06	0.1333

Fig. 7: Results showing Pooled t-test for JobSatisfaction

Variable: RelationshipSatisfaction (1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied)

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		2603	3.4391	1.1583	0.0227	1.0000	5.0000
Male		2417	3.3881	1.1662	0.0237	1.0000	5.0000
Diff (1-2)	Pooled		0.0510	1.1621	0.0328		
Diff (1-2)	Satterthwaite		0.0510		0.0328		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		3.4391	3.3946	3.4836	1.1583	1.1277	1.1907
Male		3.3881	3.3416	3.4346	1.1662	1.1343	1.2001
Diff (1-2)	Pooled	0.0510	-0.0133	0.1154	1.1621	1.1398	1.1853
Diff (1-2)	Satterthwaite	0.0510	-0.0133	0.1154			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	5018	1.55	0.1202
Satterthwaite	Unequal	4985.3	1.55	0.1203

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2416	2602	1.01	0.7329

Fig. 8: Results showing Pooled t-test for RelationshipSatisfaction

Variable: WorkLifeBalance (1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied)

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		2603	3.4068	1.1316	0.0222	1.0000	5.0000
Male		2417	3.4088	1.1680	0.0238	1.0000	5.0000
Diff (1-2)	Pooled		-0.00193	1.1493	0.0325		
Diff (1-2)	Satterthwaite		-0.00193		0.0325		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		3.4068	3.3633	3.4503	1.1316	1.1017	1.1633
Male		3.4088	3.3622	3.4554	1.1680	1.1360	1.2019
Diff (1-2)	Pooled	-0.00193	-0.0656	0.0617	1.1493	1.1273	1.1722
Diff (1-2)	Satterthwaite	-0.00193	-0.0657	0.0618			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	5018	-0.06	0.9525
Satterthwaite	Unequal	4962.5	-0.06	0.9526

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2416	2602	1.07	0.1130

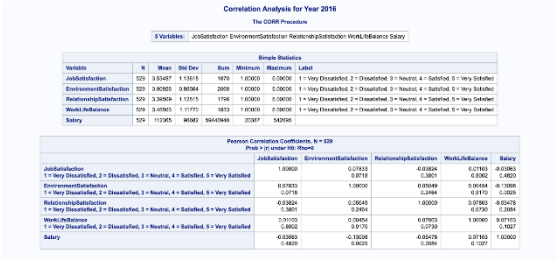
Fig. 9: Results showing Pooled t-test for WorkLifeBalance

After that the research question was tested by pooled t-test using the PROC TTEST statement to test whether there are any significant differences in performance ratings (employee's salary, work environment satisfaction, job satisfaction, workplace relationship satisfaction, and work-life balance) between male and female employees [4]. It is assumed that the significance level is at 5%, the distribution is normal, and the population standard deviation is equal and unknown.

The results show that the p-value of each variable, which is the salary of employees and all performance ratings including environment satisfaction, job satisfaction, relationship satisfaction and work-life balance is more than 0.05 (5% significance level). Therefore, there is no significant difference in employees' salary and performance ratings including environment satisfaction, job satisfaction, workplace relationship satisfaction and work-life balance between male and female employees.

To conclude, although the mean average ratings for each variable show differences between male and female employees, the t-test shows that there are no significant differences between male and female employees in terms of salary, work environment satisfaction, job satisfaction, workplace relationship satisfaction, and work-life balance. This means that there is no gender-based disparities in the workplace in terms of all the aspects mentioned above. It shows that gender does not have a huge influence on employees' salary and performance ratings. It reflects that the company has a fair workplace and equitable policies to ensure both male and female employees can receive equal treatment on salary, opportunities and training resources. Since the average rating is above 3, it draws the conclusion that both male and female employees are generally satisfied with their workplace environment, job, workplace relationship and work-life balance.

(3) Which factor—work-life balance, relationship satisfaction, environment satisfaction or salary—has the strongest impact on job satisfaction?


Correlation Analysis for Year 2016

The CORR Procedure

5 Variables: JobSatisfaction EnvironmentSatisfaction RelationshipSatisfaction WorkLifeBalance Salary

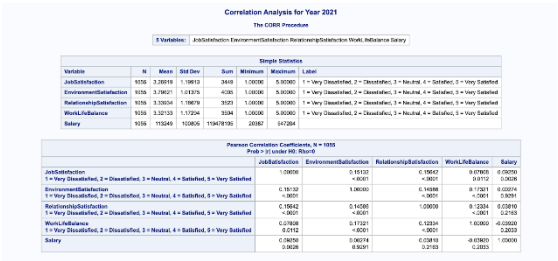
Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
JobSatisfaction	529	3.53497	1.13918	1870	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
EnvironmentSatisfaction	523	3.90805	0.98364	2044	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
RelationshipSatisfaction	529	3.39509	1.12515	1796	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
WorkLifeBalance	529	3.46965	1.11770	1835	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
Salary	529	112365	96182	59443848	20387	542695	

Pearson Correlation Coefficients, N = 529
Prob > |r| under H0: Rho=0

	JobSatisfaction	EnvironmentSatisfaction	RelationshipSatisfaction	WorkLifeBalance	Salary
JobSatisfaction 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	1.00000	0.07833 0.0718	-0.03824 0.3801	0.07163 0.1002	-0.03063 0.4820
EnvironmentSatisfaction 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	0.07833 0.0718	1.00000	0.05449 0.2484	0.00454 0.9179	-0.11695 0.0026
RelationshipSatisfaction 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	-0.03824 0.3801	0.05449 0.2484	1.00000	0.07963 0.0730	-0.00478 0.2884
WorkLifeBalance 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	0.07163 0.1002	0.00454 0.9179	0.07963 0.0730	1.00000	0.07163 0.1027
Salary	-0.03063 0.4820	-0.11695 0.0026	-0.00478 0.2884	0.07163 0.1027	1.00000

Fig. 10: Results showing Correlation Analysis for Year 2016


Correlation Analysis for Year 2021

The CORR Procedure

5 Variables: JobSatisfaction EnvironmentSatisfaction RelationshipSatisfaction WorkLifeBalance Salary

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
JobSatisfaction	1055	3.26918	1.19813	3449	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
EnvironmentSatisfaction	1055	3.79621	1.01875	4005	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
RelationshipSatisfaction	1055	3.33834	1.19679	3523	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
WorkLifeBalance	1055	3.32133	1.17294	3504	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
Salary	1055	113249	100806	119478195	20387	547264	

Pearson Correlation Coefficients, N = 1055
Prob > |r| under H0: Rho=0

	JobSatisfaction	EnvironmentSatisfaction	RelationshipSatisfaction	WorkLifeBalance	Salary
JobSatisfaction 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	1.00000	0.15132 <.0001	0.15642 <.0001	0.07805 0.0112	-0.09759 0.0015
EnvironmentSatisfaction 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	0.15132 <.0001	1.00000	0.14088 <.0001	0.17321 <.0001	0.02274 0.4351
RelationshipSatisfaction 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	0.15642 <.0001	0.14088 <.0001	1.00000	0.12334 <.0001	0.03816 0.2155
WorkLifeBalance 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	0.07808 0.0112	0.17321 <.0001	0.12334 <.0001	1.00000	-0.03020 0.3263
Salary	-0.09759 0.0015	0.02274 0.4351	0.03816 0.2155	-0.03020 0.3263	1.00000

Fig. 11: Results showing Correlation Analysis for Year 2021

In the Performance Rating dataset, the performance ratings have been recorded since 2013 and continue until 2022. For this question, we only look at data from 2016 and 2021 as these two years fall inside the range and have a gap of 5 years, which allows us to see if there are any changes in the factors over time.

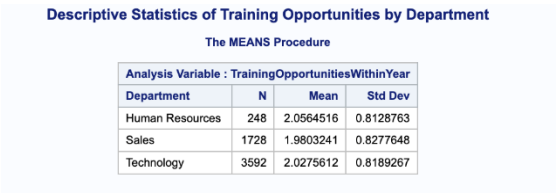
First, we filtered out data for 2016 and 2021 to make a simple descriptive analysis. The fact that there are just 529 reviews overall from the 2016 data suggests that there were still not many employees at that time. Meanwhile, there are 1055 ratings in 2021, which shows a significant increase in employees and reflects that the company has expanded its size over time. By comparing the descriptive analysis of both years, a decrease in job satisfaction is observed from the data where the mean decreased from 3.53 in 2016 to 3.27 in 2021. Out of all the variables, work-life balance and environment satisfaction have significantly declined between 2016 and 2021. However, it is still unclear whether there is a relationship between these factors and the level of job satisfaction, which will be confirmed further by doing the Pearson Correlation Analysis.

A correlation analysis is conducted to find out the relationship between job satisfaction and each factor. The Pearson correlation table for dataset 2016 shows that none of the factors have an impact on employees' job satisfaction and none of them are statistically significant. Therefore, the results might be due to random chance. On the other hand, for the year 2021, environment satisfaction and relationship satisfaction are both statistically significant ($p < 0.0001$). However, the correlations for both factors are too weak to draw any meaningful conclusions. This indicates that employees who are more satisfied with their working environment and have more positive relationships in their workplace are more likely to have slightly higher job satisfaction. The other factors, which are work-life balance and salary also show significant correlations but have a weaker impact on job satisfaction.

In short, there are no variables that show a significant correlation with job satisfaction in 2016, but by 2021, some of the variables have shown significant correlations with job satisfaction. This may be caused by the increase in the number of employees and performance ratings that bring a more accurate result. Among all the factors, environmental satisfaction stands out in both 2016 and 2021. Although the relationship is very weak, it still shows that the working environment has a minor impact on job satisfaction, and companies should focus on improving the workplace to increase the job satisfaction of their employees so that they are more willing to stay in the company. However, to answer this question more accurately, more potential factors like career development should be explored and considered as all the variables that we examined in the dataset do not show strong correlations.

(4) How does access to training opportunities vary across departments?

This analysis aimed to determine whether there are significant differences in training opportunities across the Human Resources, Sales, and Technology departments. The primary goal was to explore whether some departments have better training resources than others. These findings will help uncover whether the allocation of training opportunities is equitable across these key areas of the organization.


Descriptive Statistics of Training Opportunities by Department

The MEANS Procedure

Analysis Variable : TrainingOpportunitiesWithinYear			
Department	N	Mean	Std Dev
Human Resources	248	2.0564516	0.8128763
Sales	1728	1.9803241	0.8277648
Technology	3592	2.0275612	0.8189267

Fig. 12: Results showing Descriptive Statistics of Training Opportunities by Department

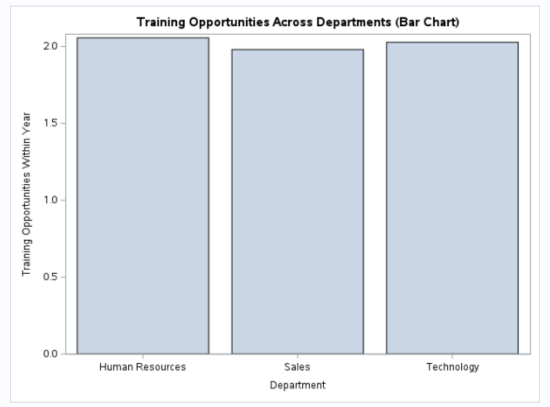


Fig. 13: Results showing Bar Chart for Training Opportunities Across Departments

The means and standard deviations for training opportunities of the three departments indicated the distribution of training opportunities is relatively balanced and consistent within each department. This evenness of distribution suggests that the company might be implementing a standardized policy for training allocation, ensuring that all employees in different departments have equal access to professional development resources. Moreover, the bar chart shows that the average training opportunities are evenly distributed across departments, with an average of 2.0 per department. To conclude, the evenly distributed training opportunities as observed from the summary statistics and bar graph reflect the organization's consistent strategy in providing professional development, ensuring each department receives equal attention.

ANOVA for Training Opportunities Across Departments

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
Department	3	Human Resources Sales Technology

Number of Observations Read	5568
Number of Observations Used	5568

ANOVA for Training Opportunities Across Departments

The ANOVA Procedure

Dependent Variable: TrainingOpportunitiesWithinYear

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3.066998	1.533499	2.27	0.1031
Error	5565	3754.812132	0.674719		
Corrected Total	5567	3757.879131			

R-Square	Coeff Var	Root MSE	TrainingOpportunitiesWithinYear Mean
0.000816	40.78134	0.821413	2.014188

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Department	2	3.06699828	1.53349914	2.27	0.1031

Fig. 14: Results showing ANOVA test for Training Opportunities Across Departments

ANOVA for Training Opportunities Across Departments

The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for TrainingOpportunitiesWithinYear

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	5565
Error Mean Square	0.674719
Critical Value of Studentized Range	3.31538

Comparisons significant at the 0.05 level are indicated by ***.

Department Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
Human Resources - Technology	0.02889	-0.09754	0.15532
Human Resources - Sales	0.07613	-0.05463	0.20689
Technology - Human Resources	-0.02889	-0.15532	0.09754
Technology - Sales	0.04724	-0.00914	0.10361
Sales - Human Resources	-0.07613	-0.20689	0.05463
Sales - Technology	-0.04724	-0.10361	0.00914

Fig. 15: Results showing Tukey's Studentized Range (HSD) Test for TrainingOpportunitiesWithinYear

Furthermore, the ANOVA test produced a p-value of 0.1031 > 0.05, indicating that there are no statistically significant differences in the training opportunities provided to employees across the three departments. This may be due to the reasons that the company is allocating a fixed number of training opportunities to each department without considering the specific needs or roles of the department. Additionally, analysis using Tukey's Honest Significant Difference (HSD) test confirmed that the differences in training opportunities between any two departments were not statistically significant. In conclusion, these findings suggest that the company prioritizes equity in its training approach, ensuring equal access to development opportunities for all employees and fostering a culture of fairness in the workplace. This ensures that every employee feels equally valued and supported in their professional growth.

CONCLUSION

In conclusion, this study explores the various factors influencing employee's performance, satisfaction, and training opportunities across departments and genders. Through data cleaning and statistical analysis by using SAS programming language, the findings demonstrated that manager's performance ratings and training opportunities are evenly distributed across departments and genders. It shows that Atlas Lab maintains a fair and balanced workplace environment with equitable performance evaluation criteria and training opportunities. However, it should be noted that the great differences in the number of employees across different departments may influence the accuracy of the results.

Other than that, by comparing the employees' job satisfaction in 2016 and 2021, it is discovered that the workplace environment has an increasing impact on employees' job satisfaction even though the

relationship is relatively weak. Therefore, to investigate the most impactful factor affecting employees' job satisfaction, more variables such as career development or benefits should be explored and analysed to answer this question better. This would be useful in helping businesses to prioritise improvements in aspects that improve job satisfaction, eventually encouraging higher productivity in employees in the future.

There are also some weaknesses of the study that should be noted. This is due to the reasons that while conducting the analysis using t-tests and ANOVA, certain assumptions such as normality and homogeneity of variance were made. However, the validity of these assumptions was not accurately tested during the study. To strengthen the reliability of the findings, additional tests such as Levene's test for homogeneity of variances should be conducted to ensure these assumptions are met. Moreover, while the study focuses on the quantity of training opportunities, it does not evaluate their quality and relevance to departmental goals. Future research should focus on addressing these limitations by exploring the specific departmental needs and assessing the impact of training on employees' performance.

ACKNOWLEDGEMENT

IST2034 Analytics Engineering lecturer and students gratefully acknowledge the datasets shared by **Mahmoud Emad Abdallah** for learning purposes.
Datasets:
https://www.kaggle.com/datasets/mahmoudemadabdallah/hr-analytics-employee-attrition-and-performance?select=Employee.csv

DATASETS

LINK:
https://drive.google.com/drive/folders/1gkQTSADJPwBFhGl109mo-AXvxqQwcneQ?usp=sharing

REFERENCES

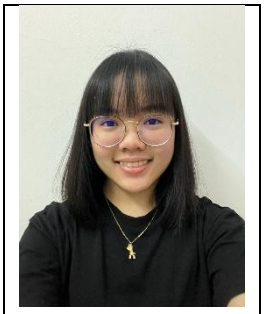
[1] M. E. Abdallah, "HR Analytics Employee Attrition & Performance." Kaggle. Accessed: Dec. 5, 2024. [Online.] Available: https://www.kaggle.com/datasets/mahmoudemadabdallah/hr-analytics-employee-attrition-and-performance

[2] SAS Institute Inc., "Base SAS ® Procedures Guide SAS ® Documentation." SAS Documentation. Accessed: Dec. 9, 2024. [Online.] Available: https://documentation.sas.com/api/collections/pgmsascdc/v_058/docsets/proc/content/proc.pdf?locale=en#nameddest=n18jyszn33umngn14czw2qfw7thc

[3] SAS Institute Inc., "ANOVA procedure: Syntax." SAS Documentation. Accessed: Dec. 12, 2024. [Online.] Available: https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/statug/statug_anova_syntax01.htm

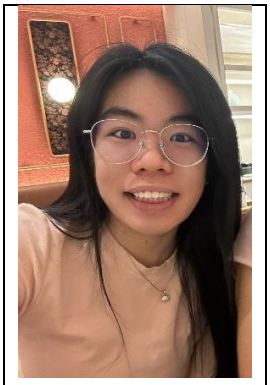
[4] SAS Institute Inc., "SAS/STAT 15.2 ® User's Guide The TTEST Procedure SAS ® Documentation." SAS Documentation. Accessed: Dec. 13, 2024. [Online.] Available: https://documentation.sas.com/api/docsets/statug/15.2/content/ttest.pdf?locale=en#nameddest=statug_ttest_syntax01

BIBLOGRAPHY



AYU WEN LI: The research question assigned to me was question 1: *How do manager ratings vary across different departments.* The purpose of this analysis was to determine if certain departments regularly received higher or lower Manager Ratings. This could help identify patterns of performance within different areas of the organization.

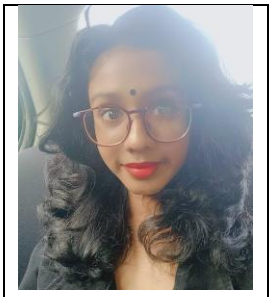
The keep=option was first used to only keep the Manager Rating and Department variables for a more efficient code. Next, descriptive statistics were then generated to analyze the variability and central tendency. Using descriptive statistics, it can be understood that employees across the three departments (Technology, Sales and Human Resources) would receive approximately similar average performance ratings from their managers in their respective departments with little variability. This indicates that on average, employees would perform higher than the benchmark, showing consistency in performance levels across departments. Furthermore, ANOVA was also conducted to examine whether the department manager's rating mean is different from the others. The findings revealed that the p-value is greater than the significance level of 0.05, indicating that there is no statistically significant difference in ratings from managers across the departments. A post hoc Tukey's studentized result (HSD) was then carried out, further confirming that there is no significant difference between the performance ratings for any pair of departments. Although the findings revealed that manager ratings showed minimal difference across the department, further analysis could be done to gain a better understanding. One such improvement would be to collect and include qualitative data such as manager's feedback to better comprehend the reasoning behind the rating.



GOO WENG XI: I am assisting Phin Han and Yan Qi with research questions 3 and 4. For research question 3 investigating the factors influencing employees' job satisfaction, the performance rating for year 2016 and 2021 was filtered to investigate if there is an increasing impact of factors (environment, relationship, work-life balance and salary) in influencing employees' job satisfaction. It was identified that although there was no correlation in both year 2016 and 2021, there is increasing correlation and statistically significant of environmental and relationship on

employee's job satisfaction increasing multicollinearity between those factors. This indicates that we should perform other tests that takes multiple factors into account such as logistic regression analysis.

On the other hand, research question 4 is studying the variations between training opportunities given to employees in different departments (Human Resource, Technology and Sales)by using ANOVA. Although the findings show no significant differences between the number of training opportunities given to employees across different departments, the number of training opportunities taken by employees is not considered. Therefore, it might not accurately reflect the effectiveness of the training program. In order to gain a clearer understanding, future studies should explore the employees' participation rates and the quality of the training program.



KEERTANA A/P SUBRAMANIAM: I played a primary role in solving research question 2 to determine if there are any significant gender disparities between male and female employees using the satisfaction ratings along with salary. Firstly, the dataset was filtered to only include male and female genders, satisfaction ratings and salary. After the dataset was properly filtered, I analyzed the descriptive statistics which revealed that there is a slight difference between male and female employee in terms of the satisfaction ratings and salary. The male employees had a higher average salary and higher environment satisfaction while female employees had a higher job and workplace relationship satisfaction compared to male employees. However, when it came to work-life balance satisfaction, both genders had similar ratings.

On the contrary, pooled t - test was also conducted to further investigate the significant differences between male and female employees. The t-test revealed that the p-values for salary and the satisfaction ratings were all greater than the significance level, which is 0.05, resulting in no significant differences between male and female employees. Hence, based on the findings, it can be concluded that gender does not significantly influence salary or performance ratings suggesting that any observed differences in satisfaction ratings and salary are not statistically meaningful.

LIM PHIN HAN: In this assessment, my main contributions focus on two parts: data cleaning and manipulation, and the third research question, which aims to find out the factors that affect employees' job satisfaction. For the data cleaning process, my part is to check for outliers for all the numerical variables. After running

the code, the result will show basic descriptive statistics and also more detailed information like quartiles and extreme observations of each variable to check for outliers. The next part of my coding is to calculate the frequency of each categorical variable to identify unusual distributions that may reflect outliers.

The second part of my contribution will be research question 3. For this question, I worked with my groupmate Goo Weng Xi as she filtered out the data from 2016 and 2021 that we chose to analyse while I generated the descriptive statistics and correlation analysis for both years. The results indicate that in 2016, the level of job satisfaction was not influenced by any of the factors, but in 2021, environment satisfaction and relationship satisfaction had a minor impact on job satisfaction. Nevertheless, the correlation is too weak and might not be precise enough to conclude that these two criteria alone have an impact on job satisfaction. Although the correlation is weak, there is still an increase in the correlation of these two factors on job satisfaction between these five years. From this result, we can conclude that additional information is needed for analysis to identify the insights that influence job satisfaction.

SIOW QI YUNG: I am responsible of doing research question 1 and 2. In research question 1, ANOVA test shows that there is no significant difference in the ratings from managers across the different departments. Since the mean for each department is above 3 from the descriptive statistics, each employee is performing equally well in different departments. This may suggest that all managers in different departments are satisfied with the performance of employees. This may show a fair workplace because every employee has equal training opportunities and other treatments. It also shows that managerial performance evaluations might be consistent regardless of which department. The company can maintain using the same evaluation practices across different departments. For research question 2, t-test shows that there is no significance difference in salary, satisfaction with environment, job, workplace relationship and work-life balance between male and female employees. It means that no gender-based disparities in those aspects. Since the average ratings for each performance ratings are above 3, employees are generally satisfied with their workplace, including the environment and relationships and they are working in a fair, inclusive workplace, and receiving fair treatment on salary, opportunities and training

resources, regardless of gender. This also shows that the company is supporting an inclusive workplace with upholds gender equality and eliminates bias on gender.

THAM YAN QI: The research question I addressed is Question 4. It examines the variation in training opportunities across three departments which is Human Resources, Sales, and Technology. A bar chart was generated using PROC SGPLOT to visualize the mean number of training opportunities each department receives. The chart highlights that the Technology department offers the highest average training opportunities, followed closely by Human Resources, with Sales presenting the lowest average.

To statistically assess these differences, an ANOVA was conducted using PROC ANOVA. It aimed to determine whether the variations in training opportunities among departments are significant. The results indicated that there is no statistically significant difference in training opportunities across the departments ($p = 0.1031$). The p-value suggest that any observed differences in means are likely due to random variation rather than departmental policies or practices. Furthermore, post-hoc comparisons using Tukey's HSD test were performed to identify specific pairs of departments with significant differences. The results confirmed that none of the pairwise comparisons reached statistical significance at the 0.05 level, reinforcing the conclusion that training opportunities are relatively consistent across Human Resources, Sales, and Technology departments.

In conclusion, visual and statistical analyses demonstrate that access to training opportunities does not significantly vary among the departments studied. This suggests a uniform approach to employee development within the organization.

APPENDIX

Group Member Name	Code Segment
Ayu Wen Li	Data Pre-Processing and Validation:

```
/*-------------------------Code below is done by Ayu Wen Li---------------------------- */

/* Saving under permanent library name 'ASSGN' */
LIBNAME ASSGN '/home/u63866143/AssignmentDataset';

/* Import the Employee.csv file into SAS */
PROC IMPORT DATAFILE='/home/u63866143/AssignmentDataset/Employee.csv'
            OUT=ASSGN.Employee_Data
            DBMS=csv
            REPLACE;
    GETNAMES=yes; /* Use the first row as variable names */
RUN;

/* Applying the correct informat and format for the HireDate column */
DATA ASSGN.Employee_Data;
    SET ASSGN.Employee_Data;

    INFORMAT HireDate mmddyy10.;  /* INFORMAT used to read the date as MM/DD/YYYY */
    FORMAT HireDate DATE9.;     /* FORMAT used to display the date as DDMONYYYY */

    /* Adding labels to Education variable */
    LABEL Education =  "1 = No Formal Qualifications,
                        2 = High School,
                        3 = Bachelors,
                        4 = Masters,
                        5 = Doctorate";

RUN;

/* Printing the first 10 observations to verify the data to ensure the dates are correctly formatted */
PROC PRINT DATA=ASSGN.employee_data (OBS = 10);
RUN;
```

Results (First 10 observations):

Research Question 1:

```
/* Question 1: How do manager ratings vary across different departments?
Objective: To analyze and compare the distribution of manager ratings (employees' performance ratings given by managers) across
various departments within the organization. */

/*Keep only relevant columns related to performance ratings(ManagerRating) and departments(Department) */
DATA ASSGN.ManagerRating_Department_Data_Q1;
   SET ASSGN.New_Cleaned_Merged_Data(KEEP= Department ManagerRating);
RUN;

/*Descriptive Statistics for Manager Rating by Department*/
proc means data=ASSGN.ManagerRating_Department_Data_Q1 mean std min max;
    class Department;
    var ManagerRating;
    TITLE "Descriptive Statistics for Manager Ratings by Department";
run;
TITLE;

/*Perform ANOVA to analyze Manager ratings by department */
proc anova data=ASSGN.ManagerRating_Department_Data_Q1 plots=none;
    TITLE "ANOVA for Employee Performance Rating by Manager Across Departments";
    class Department;  /* Classify by Department */
    model ManagerRating = Department;  /* Analyze ManagerRating by Department */
    means Department / tukey cldiff;  /* Tukey test for pairwise comparisons */
run;
TITLE;

/* Box Plot Visualisation */
proc sgplot data=ASSGN.ManagerRating_Department_Data_Q1;
    vbox ManagerRating / category=Department;
    TITLE "Distribution of Manager Ratings by Department";
run;
```

Results:

Descriptive Statistics for Manager Ratings by Department

The MEANS Procedure

Analysis Variable : ManagerRating 1 = Unacceptable, 2 = Needs Improvement, 3 = Meets Expectation, 4 = Exceeds Expectation, 5 = Above and Beyond					
Department	N Obs	Mean	Std Dev	Minimum	Maximum
Human Resources	248	3.4354839	1.0120083	2.0000000	5.0000000
Sales	1728	3.4392361	0.9725889	2.0000000	5.0000000
Technology	3592	3.4877506	0.9486189	2.0000000	5.0000000

ANOVA for Employee Performance Rating by Manager Across Departments

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
Department	3	Human Resources Sales Technology

Number of Observations Read	5568
Number of Observations Used	5568

ANOVA for Employee Performance Rating by Manager Across Departments

The ANOVA Procedure

Dependent Variable: ManagerRating 1 = Unacceptable, 2 = Needs Improvement, 3 = Meets Expectation, 4 = Exceeds Expectation, 5 = Above and Beyond

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3.061894	1.530947	1.66	0.1894
Error	5565	5118.048558	0.919685		
Corrected Total	5567	5121.110453			

R-Square	Coeff Var	Root MSE	ManagerRating Mean
0.000598	27.63403	0.959002	3.470366

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Department	2	3.06189449	1.53094724	1.66	0.1894

ANOVA for Employee Performance Rating by Manager Across Departments

The ANOVA Procedure

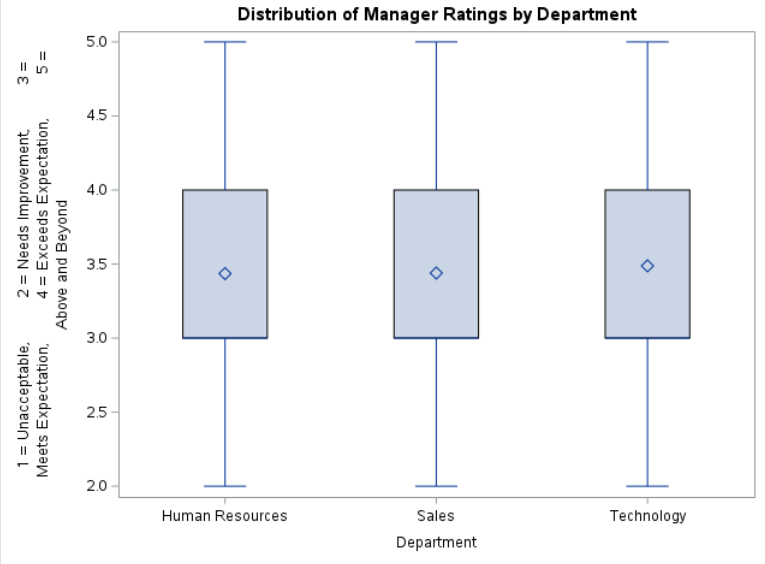
Tukey's Studentized Range (HSD) Test for ManagerRating

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	5565
Error Mean Square	0.919685
Critical Value of Studentized Range	3.31538

Comparisons significant at the 0.05 level are indicated by ***.

Department Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
Technology - Sales	0.04851	-0.01730	0.11433
Technology - Human Resources	0.05227	-0.09534	0.19987
Sales - Technology	-0.04851	-0.11433	0.01730
Sales - Human Resources	0.00375	-0.14891	0.15642
Human Resources - Technology	-0.05227	-0.19987	0.09534
Human Resources - Sales	-0.00375	-0.15642	0.14891



Goo Weng Xi	Data Pre-Processing and Validation:

```sas
/*-------------------------Code below is done by Goo Weng Xi---------------------------- */

/* This section is to check for inconsistencies in the dataset*/
/*- Duplication (duplicate id)
- Data range consistencies
- Data Integrity*/

/* Using PROC MEANS to check for valid data ranges*/
PROC MEANS DATA = ASSGN.Cleaned_Merged_Data NOLABELS;
RUN;

/* Using PROC FREQ to check for duplications and missing data*/

/*Duplicates found in the performance_data but not the employee_data.
These duplicates are due to the same employees submitting performance rating
for every years therefore we will sort and analyse by year  */

PROC FREQ DATA = ASSGN.Cleaned_Merged_Data ;
TITLE "Frequency of Duplicated Data (Employees Submitting Performance Rating Every Year)";
    TABLE EmployeeID / NOCUM;

RUN;


/*This section is to check for date integrity for review data and hire date
Review date should NOT be before hire date as review date is the date that the employees are
reviewed for their performance during their employability */
/*This section is to check for date integrity for review data and hire date
Review date should NOT be before hire date as review date is the date that the employees are
reviewed for their performance during their employability */

DATA ASSGN.Check_Date_Integrity;
    SET ASSGN.Cleaned_Merged_Data;
    IF HireDate > ReviewDate;
RUN;

/*Prints 10 observations from merged_data to check if there are any data integrity for review and hire date*/
PROC PRINT DATA=ASSGN.Check_Date_Integrity (OBS=10);
    TITLE "Data Integrity issues: Rows with HireDate After ReviewDate";
RUN;
TITLE;

/* Create a cleaned dataset with valid dates and no duplicate EmployeeIDs */
DATA ASSGN.New_Cleaned_Merged_Data;
    SET ASSGN.Cleaned_Merged_Data;
    IF HireDate < ReviewDate; /* Exclude rows with invalid dates */
RUN;

/* Print the cleaned data to verify */
PROC PRINT DATA= ASSGN.New_Cleaned_Merged_Data (OBS=10);
    TITLE "Validation for Cleaned Merged Data: Valid Dates";
RUN;
TITLE;
```

Results:

Data Integrity issues: Rows with HireDate After ReviewDate

Obs	EmployeeID	FirstName	LastName	Gender	Age	BusinessTravel	Department	DistanceFromHome (KM)	State	Ethnicity	Education	EducationField	JobRole	MaritalStatus
1	00B0-F199	Truman	Jirasek	Male	23	Some Travel	Sales	35	CA	White	1	Marketing	Sales Executive	Married
2	00B0-F199	Truman	Jirasek	Male	23	Some Travel	Sales	35	CA	White	1	Marketing	Sales Executive	Married
3	00B0-F199	Truman	Jirasek	Male	23	Some Travel	Sales	35	CA	White	1	Marketing	Sales Executive	Married
4	00B0-F199	Truman	Jirasek	Male	23	Some Travel	Sales	35	CA	White	1	Marketing	Sales Executive	Married
5	00B0-F199	Truman	Jirasek	Male	23	Some Travel	Sales	35	CA	White	1	Marketing	Sales Executive	Married
6	00B0-F199	Truman	Jirasek	Male	23	Some Travel	Sales	35	CA	White	1	Marketing	Sales Executive	Married
7	00E4-3D60	Sherilyn	Girke	Female	30	Frequent Traveller	Technology	37	CA	White	1	Business Studies	Data Scientist	Single
8	00E4-3D60	Sherilyn	Girke	Female	30	Frequent Traveller	Technology	37	CA	White	1	Business Studies	Data Scientist	Single
9	00E4-3D60	Sherilyn	Girke	Female	30	Frequent Traveller	Technology	37	CA	White	1	Business Studies	Data Scientist	Single
10	00E4-3D60	Sherilyn	Girke	Female	30	Frequent Traveller	Technology	37	CA	White	1	Business Studies	Data Scientist	Single

Validation for Cleaned Merged Data: Valid Dates

Obs	EmployeeID	FirstName	LastName	Gender	Age	BusinessTravel	Department	DistanceFromHome (KM)	State	Ethnicity	Education	EducationField	JobRole	MaritalStatus
1	005C-E0FB	Fin	O'Halleghan	Non-Binary	24	Frequent Traveller	Sales	17	CA	White	4	Marketing	Sales Executive	Married
2	005C-E0FB	Fin	O'Halleghan	Non-Binary	24	Frequent Traveller	Sales	17	CA	White	4	Marketing	Sales Executive	Married
3	005C-E0FB	Fin	O'Halleghan	Non-Binary	24	Frequent Traveller	Sales	17	CA	White	4	Marketing	Sales Executive	Married
4	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married
5	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married
6	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married
7	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married
8	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married
9	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married
10	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or	2	Computer	Machine	Married

Research Question 3 (half):

```sas
/* Question 3: Which factor—work-life balance, relationship satisfaction,
environment satisfaction or salary—has the strongest impact on job satisfaction? */

/*-------------------------Code below is done by Goo Weng Xi---------------------------- */

/*Filter the Dataset for Specific Years (2016 and 2021)*/
%macro filter_year(year);
DATA ASSGN.Filtered_Data_&year;
    SET ASSGN.new_cleaned_merged_data;
    WHERE YEAR(ReviewDate) = &year; /*filter out the performance rating in year 2016 and 2021*/
RUN;

PROC PRINT DATA=ASSGN.Filtered_Data_&year (OBS=10); /*print out first 10 observations to check*/
    TITLE "Filtered Data for Year &year";
RUN;
%mend;

/* Call the macro for 2016 and 2021 */
%filter_year(2016);  /* call the function to filter data in year 2016*/
%filter_year(2021);  /* call the function to filter data in year 2021*/

```

Results:

Filtered Data for Year 2016

Obs	EmployeeID	FirstName	LastName	Gender	Age	BusinessTravel	Department	DistanceFromHome (KM)	State	Ethnicity	Education	EducationField	JobRole	MaritalSta
1	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married
2	00D4-DD53	Joyce	Goor	Female	30	Frequent Traveller	Technology	44	CA	Black or African American	1	Computer Science	Software Engineer	Single
3	0145-DBFC	Keelia	Studde	Female	34	No Travel	Technology	8	CA	Asian or Asian American	4	Information Systems	Data Scientist	Single
4	022A-0219	Francine	Fernez	Female	32	Frequent Traveller	Technology	14	IL	American Indian or Alaska Nativ	3	Business Studies	Software Engineer	Married
5	02DA-7A72	Tadeas	Lackney	Male	29	Frequent Traveller	Sales	42	CA	White	3	Technical Degree	Sales Representative	Married
6	0322-D46B	Nikolas	Leslie	Male	32	Some Travel	Sales	39	CA	Black or African American	4	Marketing	Sales Executive	Married
7	03C5-51AD	Maxie	Banker	Female	25	Some Travel	Technology	1	IL	White	2	Computer Science	Data Scientist	Married
8	03EF-5904	Shepperd	Brittain	Male	31	Some Travel	Technology	45	IL	American Indian or Alaska Nativ	3	Other	Machine Learning Engineer	Married
9	041A-31B0	Gerri	Fullerlove	Female	34	Some Travel	Technology	41	NY	Black or African American	3	Computer Science	Software Engineer	Married
10	05B0-755F	Sollie	Arnke	Male	44	Some Travel	Technology	29	CA	Asian or Asian American	3	Computer Science	Data Scientist	Married

Filtered Data for Year 2021

Obs	EmployeeID	FirstName	LastName	Gender	Age	BusinessTravel	Department	DistanceFromHome (KM)	State	Ethnicity	Education	EducationField	JobRole	MaritalS
1	005C-E0FB	Fin	O'Halleghan	Non-Binary	24	Frequent Traveller	Sales	17	CA	White	4	Marketing	Sales Executive	Married
2	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married
3	00B0-F199	Trueman	Jirasek	Male	23	Some Travel	Sales	35	CA	White	1	Marketing	Sales Executive	Married
4	00D4-DD53	Joyce	Goor	Female	30	Frequent Traveller	Technology	44	CA	Black or African American	1	Computer Science	Software Engineer	Single
5	0210-E0D8	Yankee	Charteris	Male	36	Some Travel	Sales	7	NY	White	3	Technical Degree	Sales Executive	Single
6	022A-0219	Francine	Fernez	Female	32	Frequent Traveller	Technology	14	IL	American Indian or Alaska Nativ	3	Business Studies	Software Engineer	Married
7	0240-4D29	Kayli	Blenkiron	Non-Binary	35	Some Travel	Technology	31	CA	Asian or Asian American	4	Computer Science	Software Engineer	Single
8	02DA-7A72	Tadeas	Lackney	Male	29	Frequent Traveller	Sales	42	CA	White	3	Technical Degree	Sales Representative	Married
9	0322-D46B	Nikolas	Leslie	Male	32	Some Travel	Sales	39	CA	Black or African American	4	Marketing	Sales Executive	Married
10	03C5-51AD	Maxie	Banker	Female	25	Some Travel	Technology	1	IL	White	2	Computer Science	Data Scientist	Married

Research Question 4 (half):

```sas
/* Question 4: How does access to training opportunities vary across departments? */

/*-------------------------Code below is done by Goo Weng Xi---------------------------- */

/* Descriptive Analysis */
proc means data=ASSGN.New_Cleaned_Merged_Data n mean std nonobs;
    title "Descriptive Statistics of Training Opportunities by Department";
    class Department;  /* Group data by Department */
    var TrainingOpportunitiesWithinYear;  /* Analyze the Training Opportunities variable */
run;
```

Results:

Descriptive Statistics of Training Opportunities by Department

The MEANS Procedure

Analysis Variable : TrainingOpportunitiesWithinYear			
Department	N	Mean	Std Dev
Human Resources	248	2.0564516	0.8128763
Sales	1728	1.9803241	0.8277648
Technology	3592	2.0275612	0.8189267

Keertana A/P Subramaniam

Data Pre-Processing and Validation:

```
/*--------------------Code below is done by Keertana A/P Subramaniam --------------------*/

/* Importing the PerformanceRating.csv file into SAS */
PROC IMPORT DATAFILE='/home/u63866143/AssignmentDataset/PerformanceRating.csv'
            OUT=ASSGN.Performance_Data
            DBMS=csv
            REPLACE;
    GETNAMES=yes; /* Use the first row as variable names */
RUN;

/* Applying the correct informat and format for the ReviewDate column */
DATA ASSGN.Performance_Data;
    SET ASSGN.Performance_Data;

    INFORMAT ReviewDate mmddyy10.;  /* INFORMAT used to read the date as MM/DD/YYYY */
    FORMAT ReviewDate DATE9.;     /* FORMAT used to display the date as DDMONYYYY */

    LABEL
        EnvironmentSatisfaction =   "1 = Very Dissatisfied,
                                     2 = Dissatisfied,
                                     3 = Neutral,
                                     4 = Satisfied,
                                     5 = Very Satisfied"
        JobSatisfaction =           "1 = Very Dissatisfied,
                                     2 = Dissatisfied,
                                     3 = Neutral,
                                     4 = Satisfied,
                                     5 = Very Satisfied"
        RelationshipSatisfaction =  "1 = Very Dissatisfied,
                                     2 = Dissatisfied,
                                     3 = Neutral,
                                     4 = Satisfied,
                                     5 = Very Satisfied"
        WorkLifeBalance =   "1 = Very Dissatisfied,
                             2 = Dissatisfied,
                             3 = Neutral,
                             4 = Satisfied,
                             5 = Very Satisfied"
        SelfRating =    "1 = Unacceptable,
                         2 = Needs Improvement,
                         3 = Meets Expectation,
                         4 = Exceeds Expectation,
                         5 = Above and Beyond"
        ManagerRating = "1 = Unacceptable,
                         2 = Needs Improvement,
                         3 = Meets Expectation,
                         4 = Exceeds Expectation,
                         5 = Above and Beyond";


RUN;

/* Printing the first 10 observations to verify the data to ensure the dates are correctly formatted */
PROC PRINT DATA=ASSGN.Performance_Data (OBS=10);
RUN;
```

Results (First 10 observations):

Obs	PerformanceID	EmployeeID	ReviewDate	EnvironmentSatisfaction	JobSatisfaction	RelationshipSatisfaction	TrainingOpportunitiesWithinYear	TrainingOpportunitiesTaken	WorkLifeBalance	SelfRating	ManagerRating
1	PR01	79F7-78EC	02JAN2013	5	4	5	1	0	4	4	4
2	PR02	B61E-0F26	03JAN2013	5	4	4	1	3	4	4	3
3	PR03	F5E3-40BB	03JAN2013	3	4	5	3	2	3	5	4
4	PR04	0678-748A	04JAN2013	5	3	2	2	0	2	3	2
5	PR05	541F-3E19	04JAN2013	5	2	3	1	0	4	4	3
6	PR06	F93E-BDEF	04JAN2013	3	3	2	2	0	4	4	4
7	PR07	9E7A-1F70	08JAN2013	3	4	5	2	1	5	4	3
8	PR08	05ED-92F1	10JAN2013	4	5	4	1	1	3	3	2
9	PR09	F72D-261D	10JAN2013	4	5	2	1	1	4	5	4
10	PR10	774E-685D	11JAN2013	5	4	3	2	3	4	5	4

Research Question 2:

```
/* Saving under permanent library name 'ASSGN' */
LIBNAME ASSGN '/home/u63866143/AssignmentDataset';

/*  Research Question 2:
    Are there significant differences in performance ratings between male and female employees?
    Objective: To analyze whether there are gender-based disparities in employee performance ratings
*/

/* Filter data to include only Male and Female employees */
DATA ASSGN.Filtered_Data_Q2;
    SET ASSGN.New_Cleaned_Merged_Data;
    WHERE Gender IN ('Male', 'Female'); /* Include only Male and Female */
RUN;

/* Keep only relevant columns related to Satisfaction Ratings and Salary */
DATA ASSGN.Satisfaction_Salary_Data_Q2;
    SET ASSGN.Filtered_Data_Q2(KEEP=Gender EnvironmentSatisfaction JobSatisfaction RelationshipSatisfaction WorkLifeBalance Salary);
RUN;

/* Descriptive statistics for satisfaction variables and salary by gender */
PROC MEANS DATA=ASSGN.Satisfaction_Salary_Data_Q2 N MEAN STDDEV NOLABELS;
    CLASS Gender;
    VAR EnvironmentSatisfaction JobSatisfaction RelationshipSatisfaction WorkLifeBalance Salary;
    TITLE "Descriptive Statistics of Satisfaction Ratings and Salary by Gender";
RUN;

/* Performing t-tests to check for differences in satisfaction ratings and salary between male and female */
PROC TTEST DATA=ASSGN.Satisfaction_Salary_Data_Q2;
    CLASS Gender; /* Group data by Gender */
    VAR EnvironmentSatisfaction JobSatisfaction RelationshipSatisfaction WorkLifeBalance Salary;
    TITLE "T-Test for Satisfaction Ratings and Salary Between Male and Female Employees";
RUN;
```

Results:

Descriptive Statistics of Satisfaction Ratings and Salary by Gender

The MEANS Procedure

Gender	N Obs	Variable	N	Mean	Std Dev
Female	2603	EnvironmentSatisfaction	2603	3.8524779	0.9424956
		JobSatisfaction	2603	3.4429504	1.1762899
		RelationshipSatisfaction	2603	3.4391087	1.1583170
		WorkLifeBalance	2603	3.4068383	1.1316423
		Salary	2603	113107.48	96829.23
Male	2417	EnvironmentSatisfaction	2417	3.8622259	0.9579575
		JobSatisfaction	2417	3.4087712	1.1415059
		RelationshipSatisfaction	2417	3.3880844	1.1662289
		WorkLifeBalance	2417	3.4087712	1.1680299
		Salary	2417	116399.00	98117.79

T-Test for Satisfaction Ratings and Salary Between Male and Female Employees

The TTEST Procedure

Variable: EnvironmentSatisfaction (1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied)

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		2603	3.8525	0.9425	0.0185	1.0000	5.0000
Male		2417	3.8622	0.9580	0.0195	1.0000	5.0000
Diff (1-2)	Pooled		-0.00975	0.9500	0.0268		
Diff (1-2)	Satterthwaite		-0.00975		0.0269		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		3.8525	3.8163	3.8887	0.9425	0.9176	0.9688
Male		3.8622	3.8240	3.9004	0.9580	0.9317	0.9858
Diff (1-2)	Pooled	-0.00975	-0.0624	0.0429	0.9500	0.9317	0.9689
Diff (1-2)	Satterthwaite	-0.00975	-0.0624	0.0429			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	5018	-0.36	0.7164
Satterthwaite	Unequal	4977.3	-0.36	0.7166

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2416	2602	1.03	0.4150

Variable: JobSatisfaction (1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		2603	3.4430	1.1763	0.0231	1.0000	5.0000
Male		2417	3.4088	1.1415	0.0232	1.0000	5.0000
Diff (1-2)	Pooled		0.0342	1.1597	0.0328		
Diff (1-2)	Satterthwaite		0.0342		0.0327		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		3.4430	3.3977	3.4882	1.1763	1.1452	1.2091
Male		3.4088	3.3632	3.4543	1.1415	1.1102	1.1746
Diff (1-2)	Pooled	0.0342	-0.0300	0.0984	1.1597	1.1374	1.1828
Diff (1-2)	Satterthwaite	0.0342	-0.0300	0.0983			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	5018	1.04	0.2968
Satterthwaite	Unequal	5008.2	1.04	0.2963

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2602	2416	1.06	0.1333

Variable: RelationshipSatisfaction (1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied)

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		2603	3.4391	1.1583	0.0227	1.0000	5.0000
Male		2417	3.3881	1.1662	0.0237	1.0000	5.0000
Diff (1-2)	Pooled		0.0510	1.1621	0.0328		
Diff (1-2)	Satterthwaite		0.0510		0.0328		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		3.4391	3.3946	3.4836	1.1583	1.1277	1.1907
Male		3.3881	3.3416	3.4346	1.1662	1.1343	1.2001
Diff (1-2)	Pooled	0.0510	-0.0133	0.1154	1.1621	1.1398	1.1853
Diff (1-2)	Satterthwaite	0.0510	-0.0133	0.1154			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	5018	1.55	0.1202
Satterthwaite	Unequal	4985.3	1.55	0.1203

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2416	2602	1.01	0.7329

Variable: WorkLifeBalance (1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied)

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		2603	3.4068	1.1316	0.0222	1.0000	5.0000
Male		2417	3.4088	1.1680	0.0238	1.0000	5.0000
Diff (1-2)	Pooled		-0.00193	1.1493	0.0325		
Diff (1-2)	Satterthwaite		-0.00193		0.0325		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		3.4068	3.3633	3.4503	1.1316	1.1017	1.1633
Male		3.4088	3.3622	3.4554	1.1680	1.1360	1.2019
Diff (1-2)	Pooled	-0.00193	-0.0656	0.0617	1.1493	1.1273	1.1722
Diff (1-2)	Satterthwaite	-0.00193	-0.0657	0.0618			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	5018	-0.06	0.9525
Satterthwaite	Unequal	4962.5	-0.06	0.9526

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2416	2602	1.07	0.1130

Variable: Salary

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		2603	113107	96829.2	1897.9	22089.0	546549
Male		2417	116399	98117.8	1995.8	20387.0	547204
Diff (1-2)	Pooled		-3291.5	97451.8	2752.7		
Diff (1-2)	Satterthwaite		-3291.5		2754.1		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		113107	109386	116829	96829.2	94268.7	99533.8
Male		116399	112485	120313	98117.8	95427.8	100965
Diff (1-2)	Pooled	-3291.5	-8888.1	2105.1	97451.8	95582.1	99396.6
Diff (1-2)	Satterthwaite	-3291.5	-8890.8	2107.7			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	5018	-1.20	0.2319
Satterthwaite	Unequal	4980	-1.20	0.2321

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2416	2602	1.03	0.5078

Lim Phin Han

Data Pre-Processing and Validation:

```
/*-------------------------Code below is done by Lim Phin Han--------------------------- */

/*This section is to check the outliers*/
/*Checking for outliers for numerical data
question: need to check age, distance from home?*/
PROC UNIVARIATE DATA=ASSGN.New_Cleaned_Merged_Data;
    VAR Age
        Salary
        YearsAtCompany
        YearsInMostRecentRole
        YearsSinceLastPromotion
        YearsWithCurrManager
        TrainingOpportunitiesWithinYear
        TrainingOpportunitiesTaken
        ;
    TITLE "Outlier Detection for Numerical Variables in Cleaned Merged Data";
RUN;
TITLE;

/*Checking for outlier for categorical data*/
PROC FREQ DATA=ASSGN.New_Cleaned_Merged_Data;
    TABLES  Education
            EnvironmentSatisfaction
            JobSatisfaction
            RelationshipSatisfaction
            WorkLifeBalance
            SelfRating
            ManagerRating / MISSING;
    TITLE "Frequency Analysis for Categorical Variables in Cleaned Merged Data";
RUN;
TITLE;
```

Results:

Descriptive Analysis for Year 2016

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
JobSatisfaction	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	367	3.4986376	1.1472816	1.0000000	5.0000000
EnvironmentSatisfaction	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	367	3.9346049	0.8402942	1.0000000	5.0000000
RelationshipSatisfaction	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	367	3.4005450	1.1285174	1.0000000	5.0000000
WorkLifeBalance	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	367	3.4741144	1.1350165	1.0000000	5.0000000
Salary		367	125442.70	97000.29	20650.00	542695.00

Descriptive Analysis for Year 2021

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
JobSatisfaction	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	1019	3.2708538	1.2038104	1.0000000	5.0000000
EnvironmentSatisfaction	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	1019	3.8076546	1.0040685	1.0000000	5.0000000
RelationshipSatisfaction	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	1019	3.3444553	1.1877304	1.0000000	5.0000000
WorkLifeBalance	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	1019	3.3159961	1.1744870	1.0000000	5.0000000
Salary		1019	114583.67	101272.00	20387.00	547204.00

Correlation Analysis for Year 2016

The CORR Procedure

5 Variables:	JobSatisfaction EnvironmentSatisfaction RelationshipSatisfaction WorkLifeBalance Salary

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
JobSatisfaction	367	3.49864	1.14728	1284	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
EnvironmentSatisfaction	367	3.93460	0.84029	1444	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
RelationshipSatisfaction	367	3.40054	1.12852	1248	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
WorkLifeBalance	367	3.47411	1.13502	1275	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
Salary	367	125443	97000	46037472	20650	542695	

Pearson Correlation Coefficients, N = 367
Prob > |r| under H0: Rho=0

	JobSatisfaction	EnvironmentSatisfaction	RelationshipSatisfaction	WorkLifeBalance	Salary
JobSatisfaction 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	1.00000	0.09060 0.0830	-0.02174 0.6781	0.02987 0.5684	-0.03267 0.5326
EnvironmentSatisfaction 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	0.09060 0.0830	1.00000	0.05651 0.2803	0.02973 0.5702	-0.18296 0.0004
RelationshipSatisfaction 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	-0.02174 0.6781	0.05651 0.2803	1.00000	0.11370 0.0294	-0.06558 0.2101
WorkLifeBalance 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	0.02987 0.5684	0.02973 0.5702	0.11370 0.0294	1.00000	0.08114 0.1207
Salary	-0.03267 0.5326	-0.18296 0.0004	-0.06558 0.2101	0.08114 0.1207	1.00000

Correlation Analysis for Year 2021

The CORR Procedure

5 Variables:	JobSatisfaction EnvironmentSatisfaction RelationshipSatisfaction WorkLifeBalance Salary

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
JobSatisfaction	1019	3.27085	1.20381	3333	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
EnvironmentSatisfaction	1019	3.80765	1.00407	3880	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
RelationshipSatisfaction	1019	3.34446	1.18773	3408	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
WorkLifeBalance	1019	3.31600	1.17449	3379	1.00000	5.00000	1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied
Salary	1019	114584	101272	116760761	20387	547204	

Pearson Correlation Coefficients, N = 1019
Prob > |r| under H0: Rho=0

	JobSatisfaction	EnvironmentSatisfaction	RelationshipSatisfaction	WorkLifeBalance	Salary
JobSatisfaction 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	1.00000	0.14554 <.0001	0.15316 <.0001	0.07906 0.0116	0.08908 0.0044
EnvironmentSatisfaction 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	0.14554 <.0001	1.00000	0.14375 <.0001	0.16571 <.0001	-0.00459 0.8837
RelationshipSatisfaction 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	0.15316 <.0001	0.14375 <.0001	1.00000	0.12540 <.0001	0.03566 0.2555
WorkLifeBalance 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neutral, 4 = Satisfied, 5 = Very Satisfied	0.07906 0.0116	0.16571 <.0001	0.12540 <.0001	1.00000	-0.03535 0.2595
Salary	0.08908 0.0044	-0.00459 0.8837	0.03566 0.2555	-0.03535 0.2595	1.00000

Siow Qi Yung

Data Pre-Processing and Validation:

```
/*-------------------------Code below is done by Siow Qi Yung-------------------------- */

/* Printing the first 10 observations to verify the data to ensure the dates are correctly formatted */
PROC PRINT DATA=ASSGN.Performance_Data (OBS=10);
RUN;

/* Sorting the Employee_Data */
PROC SORT DATA=ASSGN.Employee_Data OUT=ASSGN.Employee_SortedData;
    BY EmployeeID;
RUN;

/* Sorting the Performance_Data */
PROC SORT DATA=ASSGN.Performance_Data OUT=ASSGN.Performance_SortedData;
    BY EmployeeID;
RUN;

/* Merging the datasets using the BY statement and MERGENOBY=ERROR */

OPTION MERGENOBY=ERROR;

DATA ASSGN.Merged_Data;
    MERGE ASSGN.Employee_SortedData ASSGN.Performance_SortedData;
    BY EmployeeID;
RUN;

/* Printing the first 10 observations to verify the merged data to
   ensure the data are correctly merged */
PROC PRINT DATA=ASSGN.Merged_Data (OBS=10);
RUN;
```

Results:

Obs	EmployeeID	FirstName	LastName	Gender	Age	BusinessTravel	Department	DistanceFromHome (KM)	State	Ethnicity	Education	EducationField	JobRole	MaritalStatus	Salary	StockOptionLevel	OverTime	HireDate	Attrition	YearsAtCompany	YearsInMostRecentRole
1	001A-BF88	Christy	Jurnel	Male	22	Some Travel	Technology	40	CA	White	4	Information Systems	Software Engineer	Married	27763	0	No	05SEP2021	No	1	0
2	005C-E0FB	Fin	O'Hallaghan	Non-Binary	24	Frequent Traveller	Sales	17	CA	White	4	Marketing	Sales Executive	Married	56155	1	No	26AUG2017	No	5	2
3	005C-E0FB	Fin	O'Hallaghan	Non-Binary	24	Frequent Traveller	Sales	17	CA	White	4	Marketing	Sales Executive	Married	56155	1	No	26AUG2017	No	5	2
4	005C-E0FB	Fin	O'Hallaghan	Non-Binary	24	Frequent Traveller	Sales	17	CA	White	4	Marketing	Sales Executive	Married	56155	1	No	26AUG2017	No	5	2
5	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married	126238	0	No	08MAR2012	No	10	3
6	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married	126238	0	No	08MAR2012	No	10	3
7	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married	126238	0	No	08MAR2012	No	10	3
8	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married	126238	0	No	08MAR2012	No	10	3
9	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married	126238	0	No	08MAR2012	No	10	3
10	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married	126238	0	No	08MAR2012	No	10	3

YearsSinceLastPromotion	YearsWithCurrManager	PerformanceID	ReviewDate	EnvironmentSatisfaction	JobSatisfaction	RelationshipSatisfaction	TrainingOpportunitiesWithinYear	TrainingOpportunitiesTaken	WorkLifeBalance	SelfRating	ManagerRating
1	0										
2	0	PR4867	17JUN2020	3	3	2	1	2	2	4	3
2	0	PR5870	17JUN2021	4	4	5	1	1	5	3	3
2	0	PR6165	17JUN2022	3	4	5	3	0	4	4	4
6	6	PR1165	15JUN2016	4	5	4	2	2	5	4	3
6	6	PR1731	15JUN2017	5	2	5	3	0	5	5	4
6	6	PR2395	15JUN2018	5	2	4	1	0	5	3	3
6	6	PR3168	15JUN2019	4	5	3	1	1	3	5	5
6	6	PR4068	18JUN2020	5	4	2	3	0	2	3	2
6	6	PR5871	18JUN2021	3	2	2	3	0	4	3	3

Tham Yan Qi

Data Pre-Processing and Validation:

```
/*----------------------Code below is done by Tham Yan Qi ------------------------------*/

/* Check for missing values */

TITLE "Count of Missing Values";
PROC MEANS DATA=ASSGN.Merged_Data N NMISS NOLABELS ;
RUN;
TITLE;

/* Show data with missing values */
DATA ASSGN.MissingRows;
    SET ASSGN.Merged_Data;
    IF cmiss(of _all_) > 0; /* Keep only rows with missing values */
RUN;

TITLE "Data with Missing Values";
PROC PRINT DATA = ASSGN.MissingRows (OBS=10);
RUN;
TITLE;

/* Remove Missing Values and create a new dataset */
DATA ASSGN.Cleaned_Merged_Data;
    SET ASSGN.Merged_Data;
    IF cmiss(of _all_) = 0; /*no missing values*/;
RUN;

PROC PRINT DATA=ASSGN.Cleaned_Merged_Data (OBS=10);
TITLE "Merged Data (Cleaned) with No Missing Values";
RUN;
TITLE;
```

Result:

Count of Missing Values

The MEANS Procedure

Variable	N	N Miss
Age	6899	0
DistanceFromHome (KM)	6899	0
Education	6899	0
Salary	6899	0
StockOptionLevel	6899	0
HireDate	6899	0
YearsAtCompany	6899	0
YearsInMostRecentRole	6899	0
YearsSinceLastPromotion	6899	0
YearsWithCurrManager	6899	0
ReviewDate	6709	190
EnvironmentSatisfaction	6709	190
JobSatisfaction	6709	190
RelationshipSatisfaction	6709	190
TrainingOpportunitiesWithinYear	6709	190
TrainingOpportunitiesTaken	6709	190
WorkLifeBalance	6709	190
SelfRating	6709	190
ManagerRating	6709	190

Data with Missing Values

Obs	EmployeeID	FirstName	LastName	Gender	Age	BusinessTravel	Department	DistanceFromHome (KM)	State	Ethnicity	Education	EducationField	JobRole	MaritalStatus	Salary	StockOptionLevel	OverTime	HireDate	Attrition	YearsAtCompany
1	001A-BF88	Christy	Jumel	Male	22	Some Travel	Technology	40	CA	White	4	Information Systems	Software Engineer	Married	27763	0	No	05SEP2021	No	1
2	018B-DF1E	Otha	Sopper	Female	19	Some Travel	Sales	1	CA	White	1	Economics	Sales Representative	Single	24499	0	Yes	17OCT2032	No	0
3	0375-2F04	Lani	Raddenbury	Female	19	No Travel	Sales	38	IL	White	4	Technical Degree	Sales Executive	Single	51673	2	No	30NOV2021	No	1
4	03D5-622D	Aurora	Whate	Female	21	No Travel	Sales	43	CA	White	1	Marketing	Sales Executive	Married	54355	1	Yes	31DEC2021	No	1
5	04D3-B8D3	Ann	Sivorn	Female	24	Frequent Traveller	Technology	23	NY	White	4	Computer Science	Senior Software Engineer	Single	158437	1	No	03OCT2032	No	0
6	054A-0B94	Temp	Weyland	Male	20	No Travel	Technology	11	IL	White	3	Computer Science	Data Scientist	Married	28702	2	No	14NOV2021	No	1
7	060D-02AB	Lynnet	Foord	Female	35	Some Travel	Sales	36	CA	White	2	Marketing	Manager	Married	487746	1	No	12JAN2032	No	0
8	06B1-2653	Montgomery	Gallford	Male	19	Frequent Traveller	Technology	26	CA	Black or African American	3	Information Systems	Software Engineer	Single	27474	0	No	08DEC2032	No	0
9	078A-E0F5	Antonino	Gillorist	Male	18	Some Travel	Sales	18	CA	White	4	Marketing	Sales Executive	Single	52198	0	No	11FEB2032	No	0
10	078D-B654	Cob	Cruikshank	Male	24	Some Travel	Technology	1	CA	White	4	Computer Science	Senior Software Engineer	Single	113262	0	No	30JUN2022	No	0

Merged Data (Cleaned) with No Missing Values

Obs	EmployeeID	FirstName	LastName	Gender	Age	BusinessTravel	Department	DistanceFromHome (KM)	State	Ethnicity	Education	EducationField	JobRole	MaritalStatus	Salary	StockOptionLevel	OverTime	HireDate	Attrition	YearsAtCompany
1	005C-E0FB	Fin	O'Halleghan	Non-Binary	24	Frequent Traveller	Sales	17	CA	White	4	Marketing	Sales Executive	Married	56155	1	No	26AUG2017	No	5
2	005C-E0FB	Fin	O'Halleghan	Non-Binary	24	Frequent Traveller	Sales	17	CA	White	4	Marketing	Sales Executive	Married	56155	1	No	26AUG2017	No	5
3	005C-E0FB	Fin	O'Halleghan	Non-Binary	24	Frequent Traveller	Sales	17	CA	White	4	Marketing	Sales Executive	Married	56155	1	No	26AUG2017	No	5
4	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married	126238	0	No	08MAR2012	No	10
5	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married	126238	0	No	08MAR2012	No	10
6	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married	126238	0	No	08MAR2012	No	10
7	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married	126238	0	No	08MAR2012	No	10
8	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married	126238	0	No	08MAR2012	No	10
9	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married	126238	0	No	08MAR2012	No	10
10	00A3-2445	Wyatt	Ziehm	Male	30	Some Travel	Technology	6	CA	Black or African American	2	Computer Science	Machine Learning Engineer	Married	126238	0	No	08MAR2012	No	10

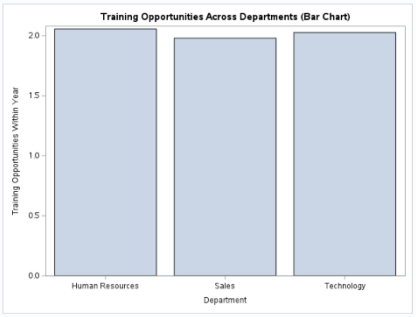
Research Question 4 (Half):

```
/*-------------------------Code below is done by Tham Yan Qi---------------------------- */

/* Bar Chart */
proc sgplot data=ASSGN.New_Cleaned_Merged_Data;
    title "Training Opportunities Across Departments (Bar Chart)";
    vbar Department / response=TrainingOpportunitiesWithinYear stat=mean; /* Bar chart showing mean training opportunities */
    xaxis label="Department";
    yaxis label="Training Opportunities Within Year";
run;

/* ANOVA */
proc anova data=ASSGN.New_Cleaned_Merged_Data plots=none;
    title "ANOVA for Training Opportunities Across Departments";
    class Department;  /* Classify by Department */
    model TrainingOpportunitiesWithinYear = Department; /* Test training opportunities across departments */
    means Department / tukey cldiff;  /* Tukey test for pairwise comparisons */
run;
```

Result:



ANOVA for Training Opportunities Across Departments

The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for TrainingOpportunitiesWithinYear

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	5565
Error Mean Square	0.674719
Critical Value of Studentized Range	3.31538

Comparisons significant at the 0.05 level are indicated by *.**

Department Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
Human Resources - Technology	0.02889	-0.09754	0.15532
Human Resources - Sales	0.07613	-0.05463	0.20689
Technology - Human Resources	-0.02889	-0.15532	0.09754
Technology - Sales	0.04724	-0.00914	0.10361
Sales - Human Resources	-0.07613	-0.20689	0.05463
Sales - Technology	-0.04724	-0.10361	0.00914

ANOVA for Training Opportunities Across Departments

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
Department	3	Human Resources Sales Technology

Number of Observations Read	5568
Number of Observations Used	5568

ANOVA for Training Opportunities Across Departments

The ANOVA Procedure

Dependent Variable: TrainingOpportunitiesWithinYear

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3.066998	1.533499	2.27	0.1031
Error	5565	3754.812132	0.674719		
Corrected Total	5567	3757.879131			

R-Square	Coeff Var	Root MSE	TrainingOpportunitiesWithinYear Mean
0.000816	40.78134	0.821413	2.014188

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Department	2	3.06699828	1.53349914	2.27	0.1031