Jordan University of Science and Technology.
Department of Computer Information Systems.
Case Study -Tableau-Dec 2023.

By
Nooralhuda Jarardat
Jelan khweileh

Supervised by
Dr. Mostafa Ali

Table of Contents:

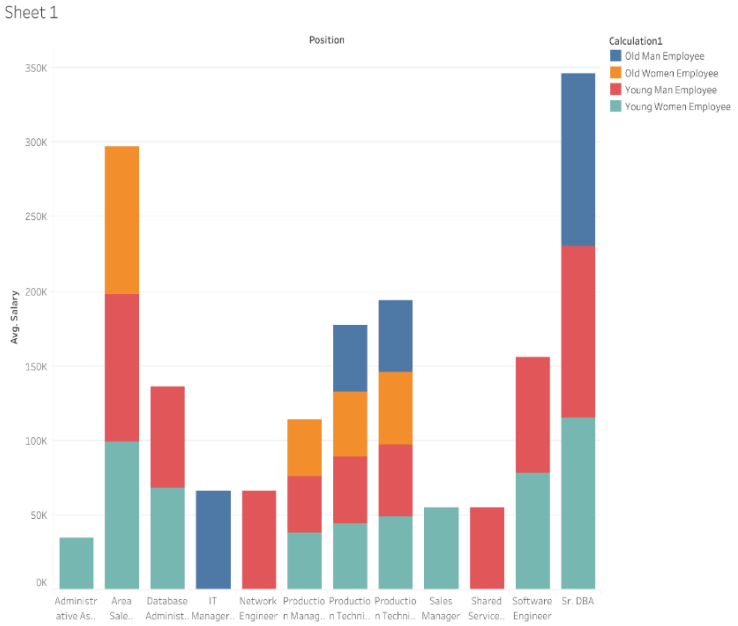
classification Section: The dataset used for this analysis is an HR dataset that includes information about employees, such as age, gender, salary, position. The objective is to create a visualization of how employees are distributed across four categories and understand the relationship between salary and position within each group.

1-Create a New Calculated Field for Classification:

Calculation1

```
IF [Sex] = 'Male' AND [Age] <= 50 THEN 'Young Man Employee'
ELSEIF [Sex] = 'Male' AND [Age] > 50 THEN 'Old Man Employee'
ELSEIF [Sex] = 'Female' AND [Age] <= 50 THEN 'Young Women Employee'
ELSEIF [Sex] = 'Female' AND [Age] > 50 THEN 'Old Women Employee'
END
```

2-Relationship between average salary and position based on "calculation1"(gender and age) as a color filter.



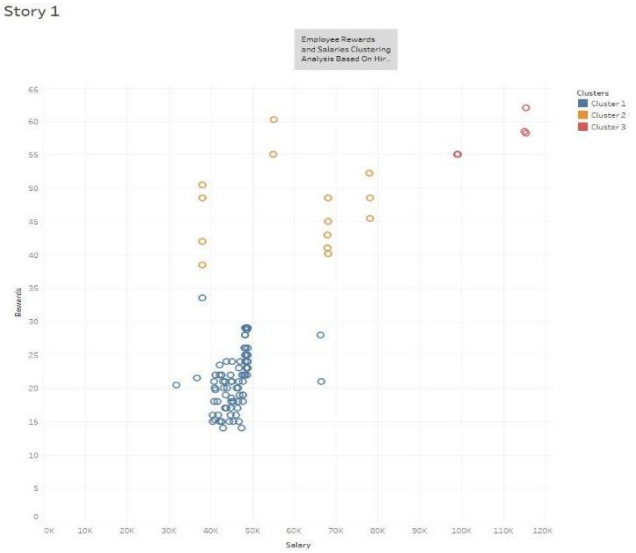
The X-axis shows the different positions in the company, and the Y-axis shows the average salaries of employees in each position.

3-Description and Findings:

- In most cases, the average salary of an "old man employee" or "old woman employee" is higher than the salary of a "young man employee" or "young woman employee," regardless of the position.
- The "old man employee" get the highest average salary in the Sr.DBA position.
- Only "young woman employee" occupy some positions, such as sales manager and administrative assistant.
- Only "young man employee" occupy some positions, such as network engineer and shared services manager.
- Only "old man employee" occupy IT manager position.
- The administrative assistant gets the lowest average salary in the company.

Clustering Section: The goal of clustering is to identify patterns to better understand relationship between rewards, salaries, and hiring date, similar data will be grouped together in one cluster.

1- Tableau Visualization:



2- Description and Findings:

After performing the clustering analysis based on salaries, rewards and hiring date, we observed interesting patterns within the dataset. The three clusters have distinct characteristics, offering information on the organization's diverse staff profiles.
More details of the clusters are shown below:

Describe Clusters

Summary | Models

Inputs for Clustering

Variables:	Sum of Rewards
	Sum of Salary
Level of Detail:	Year of Date of Hire
Scaling:	Normalized

Summary Diagnostics

Number of Clusters:	3
Number of Points:	103
Between-group Sum of Squares:	9.1178
Within-group Sum of Squares:	1.6998
Total Sum of Squares:	10.818

Centers

Clusters	Number of Items	Sum of Rewards	Sum of Salary
Cluster 1	83	21.437	45934.0
Cluster 2	14	47.037	59798.0
Cluster 3	6	57.283	1.0718e+05
Not Clustered	0		

Describe Clusters

Summary | Models

Analysis of Variance:

			Model		Error	
Variable	F-statistic	p-value	Sum of Squares	DF	Sum of Squares	DF
Sum of Rewards	43.31	2.842e-14	5.943	2	6.862	100
Sum of Salary	40.13	1.604e-13	3.175	2	3.956	100

3-Cluster Characteristics:

- ✓ Cluster 1 has the largest number of items and is characterized by a medium centroid for both rewards and salaries.
- ✓ Cluster 2 has a smaller number of items with higher centroid values for rewards and salaries.
- ✓ Cluster 3 the smallest cluster, stands out with the highest centroid values for both rewards and salaries.