PROJECT REPORT

ON

EMPLOYEE ATTRITION

Submitted By:

Sherin Joseph
09 / NOV / 2021

INTRODUCTION

- The project deals with the employee attrition in a company.
- The company uses the observational data to analyse and interpret departments in which highest attrition occurs.
- They would like to know which category of employees are more subjected to attrition.
- Moreover, the company also tries to figure out the reasons for attrition.
- Tools such as tableau prep and tableau desktop are used for analysis purpose.

TABLEAU PREP

- The employee data is distributed across 4 different files:
    1. General_data
    2. Employee_survey_data
    3. Manager_survey_data
    4. AvgSecondsworked

The tableau prep is used to eliminate unwanted columns, fill missing values, combine the different files, and generate single file output.

Familiarizing Data Set

General_data

general_data 25 fields | Filter Values...

Clear the check box to remove fields. You can also filter your data or change data types. Add a clean step to view and clean data.

Fields selected: 21 of 25

Type	Field Name	Original Field Name	Changes	Preview
#	Emp_ID	Emp_ID		1, 2, 3
#	Age	Age		51, 31, 32
Abc	Attrition	Attrition		No, Yes
Abc	BusinessTravel	BusinessTravel		Travel_Rarely, Travel_Frequently
Abc	Department	Department		Sales, Research & Development
#	DistanceFromH...	DistanceFromHome		6, 10, 17
#	Education	Education		2, 1, 4
Abc	EducationField	EducationField		Life Sciences, Other
#	EmployeeCount	EmployeeCount		1
#	EmployeeID	EmployeeID		1, 2, 3
Abc	Gender	Gender		Female, Male
#	JobLevel	JobLevel		1, 4
Abc	JobRole	JobRole		Healthcare Representative, Research Scientist, Sales Executive
Abc	MaritalStatus	MaritalStatus		Married, Single
#	AnnualIncome	AnnualIncome		131,160, 41,890, 193,280
#	NumCompanies...	NumCompaniesWorked		1, 0
Abc	Over18	Over18		Y
#	PercentSalaryHi...	PercentSalaryHike		11, 23, 15
#	StandardHours	StandardHours		8

Employee_survey_data

employee_survey_data 4 fields | Filter Values...

Clear the check box to remove fields. You can also filter your data or change data types. Add a clean step to view and clean data.

Fields selected: 4 of 4

Type	Field Name	Original Field Name	Changes	
#	EmployeeID	EmployeeID		1, 2, 3
#	EnvironmentSat...	EnvironmentSatisfaction		3, 2
#	JobSatisfaction	JobSatisfaction		4, 2
#	WorkLifeBalance	WorkLifeBalance		2, 4, 1

Manager_survey_data

manager_survey_data 3 fields | Filter Values...

Clear the check box to remove fields. You can also filter your data or change data types. Add a clean step to view and clean data.

Fields selected: 3 of 3

Type	Field Name	Original Field Name	Changes	
#	EmployeeID	EmployeeID		1, 2, 3
#	JobInvolvement	JobInvolvement		3, 2
#	PerformanceRat...	PerformanceRating		3, 4

AvgSecondsworked

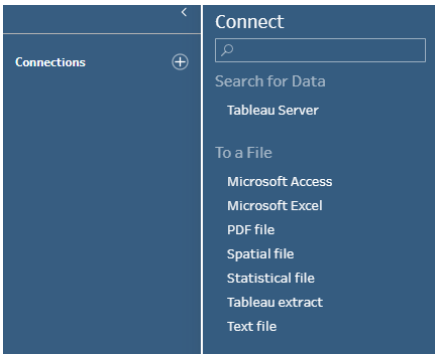
AvgSecondsworked 2 fields | Filter Values...

Clear the check box to remove fields. You can also filter your data or change data types. Add a clean step to view and clean data.

Fields selected: 2 of 2

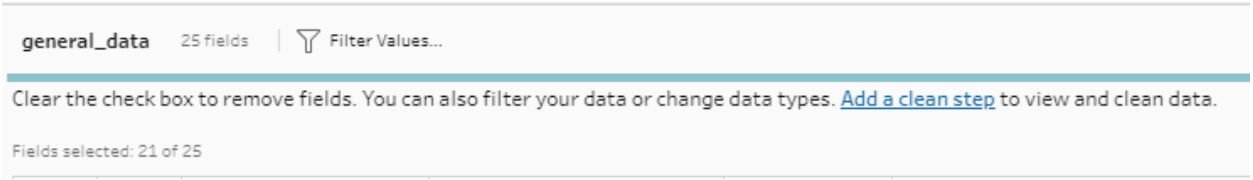
Type	Field Name	Original Field Name	Changes	
#	EmployeeID	EmployeeID		1, 2, 3
#	AvgSeconds	AvgSeconds		23,595.68, 25,126.57, 23,409.71

Loading Data to Tableau Prep:

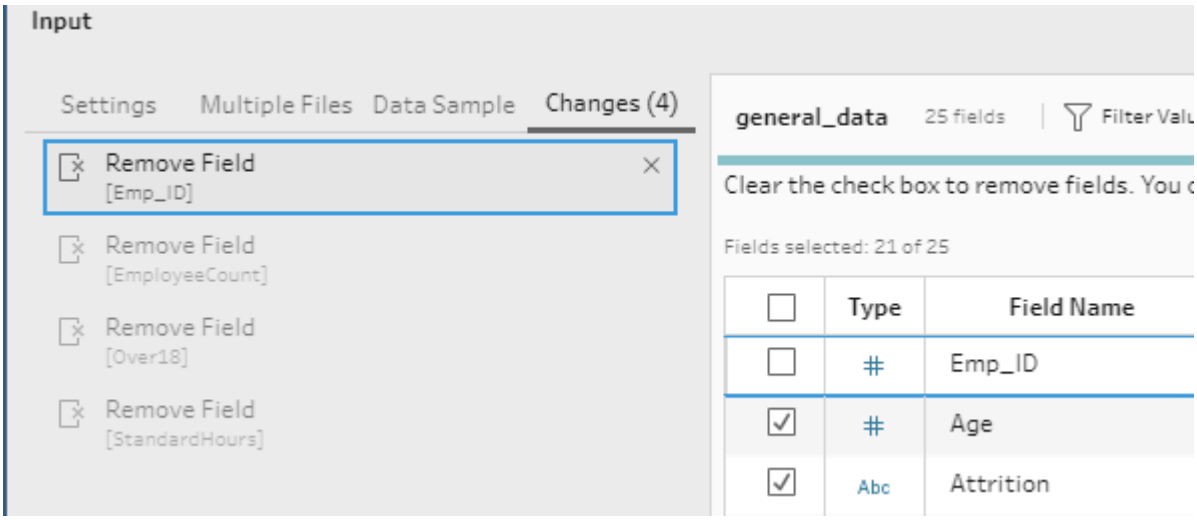


- First step is to open the tableau prep and click on connection followed by the text file to choose the desired csv file.
- Then browse the file “General_data” to get the data loaded.

Input Step:

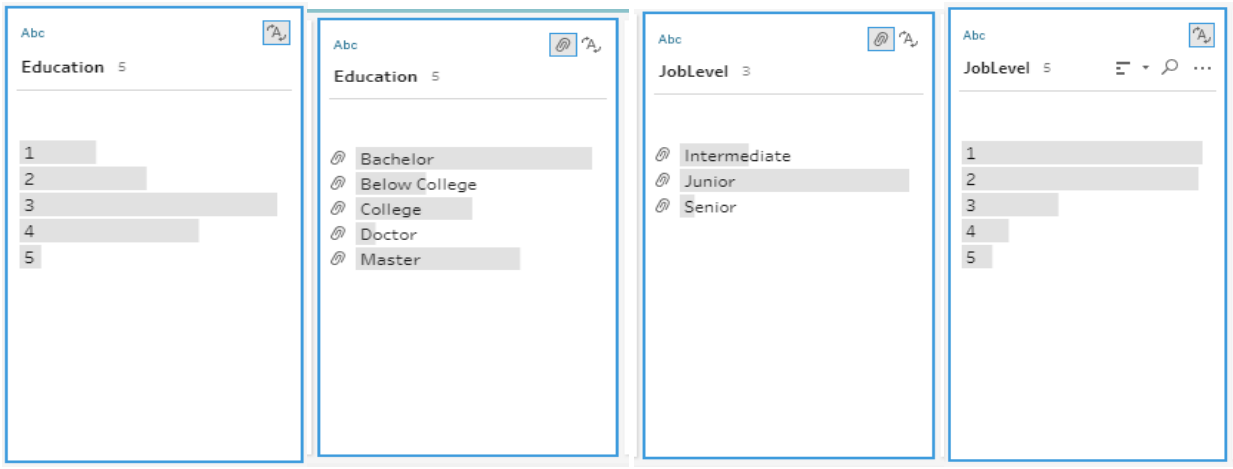


- Click on the “Add a clean step” to add the cleaning step for data cleaning.



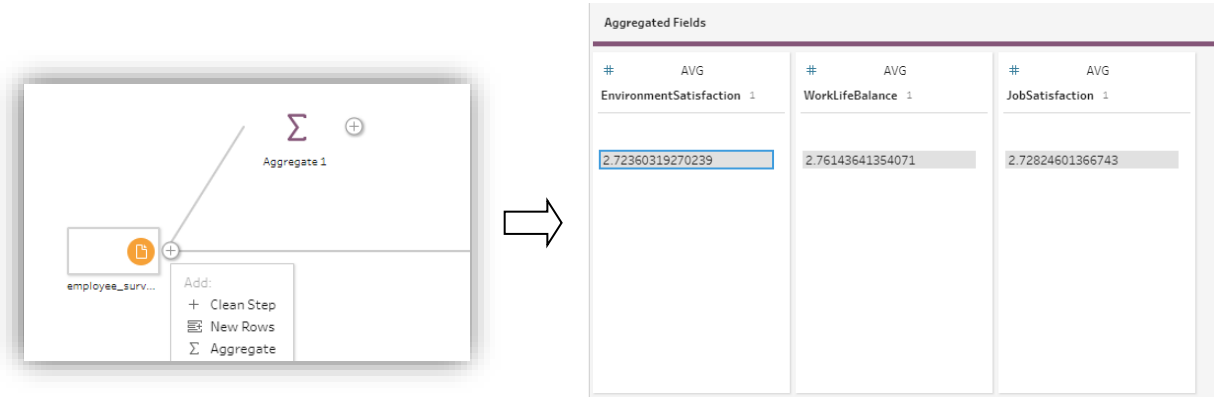
- We could see total of 4 changes in the input step.
- We have omitted the duplicate column “Emp_ID”.
- Also removed the nonrelevant columns such as “EmployeeCount”, “Over18” and “StandardHours”.

Cleaning Step: General_data



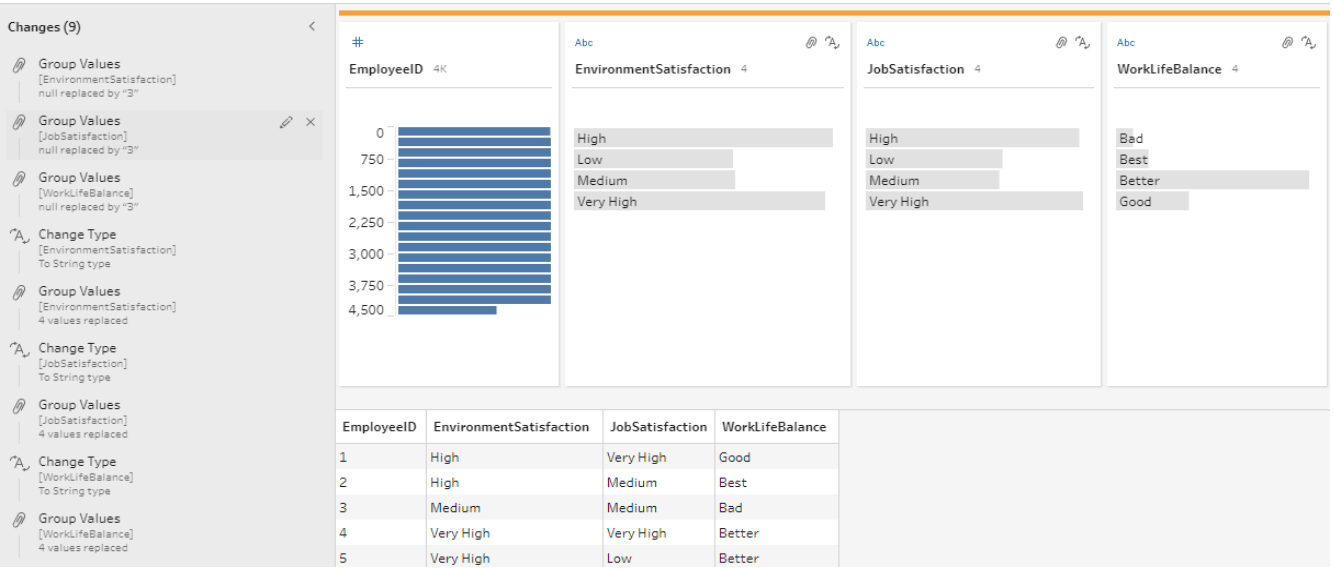
- The data type of the fields “Education” and “JobLevel” are changed from integer to String.
- The corresponding values are then mapped to the categorical values in their respective columns.
- The summary of the changes are shown in the changes pane and there are total 4 changes made.
- The cleaned output is saved in the name “CL.General.Data”

Aggregate of Employee_Survey_Data:



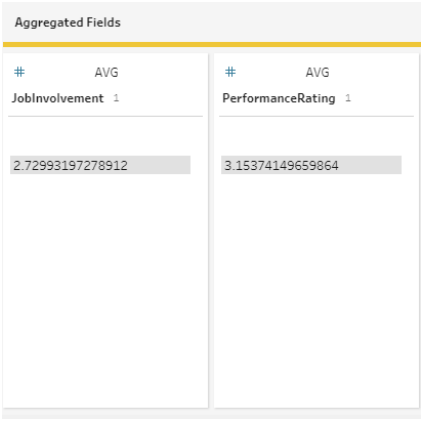
- The Aggregate step is used to find out the mean of the variables.
- These mean values are used to replace the missing values in column in the next cleaning step.

Cleaning Step: Employee_Survey_Data



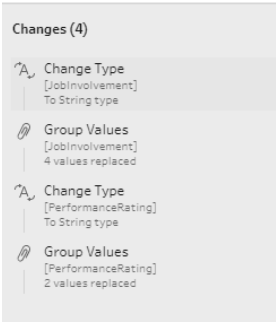
- The first step is replacing the missing values in each column with the mean value (rounded).
- Then the data type of the columns is changed from Integer to String.
- The corresponding values are then mapped to the categorical values in their respective columns.
- The summary of the changes is shown in the changes pane and there is total 9 changes made.
- The cleaned output is saved in the name “CL.Emp.Survey”

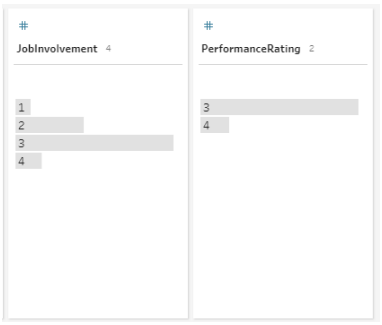
Aggregate of Manager_survey_data:

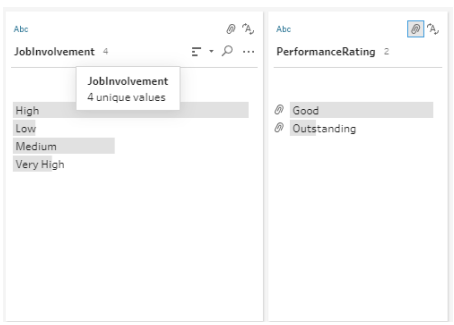


- The Aggregate step is used to find out the mean of the variables.
- These mean values are used to replace the missing values in column in the next cleaning step.

Cleaning Step: Manager_survey_data

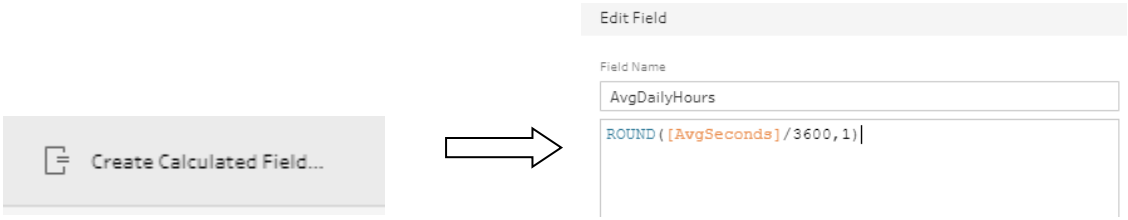




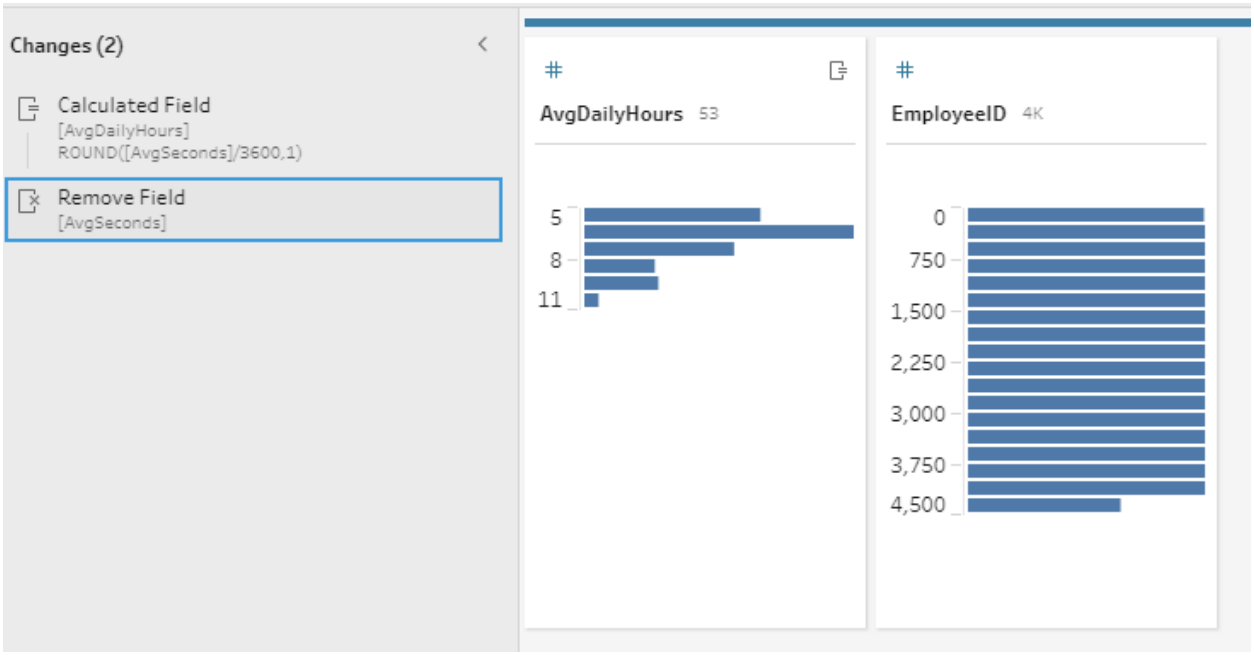


- The first step is replacing the missing values in each column with the mean value (rounded).
- Then the data type of the columns is changed from Integer to String.
- The corresponding values are then mapped to the categorical values in their respective columns.
- The summary of the changes is shown in the changes pane and there is total 4 changes made.
- The cleaned output is saved in the name “CL.Mgr.Survey”

Cleaning Step: AvgSecondsworked

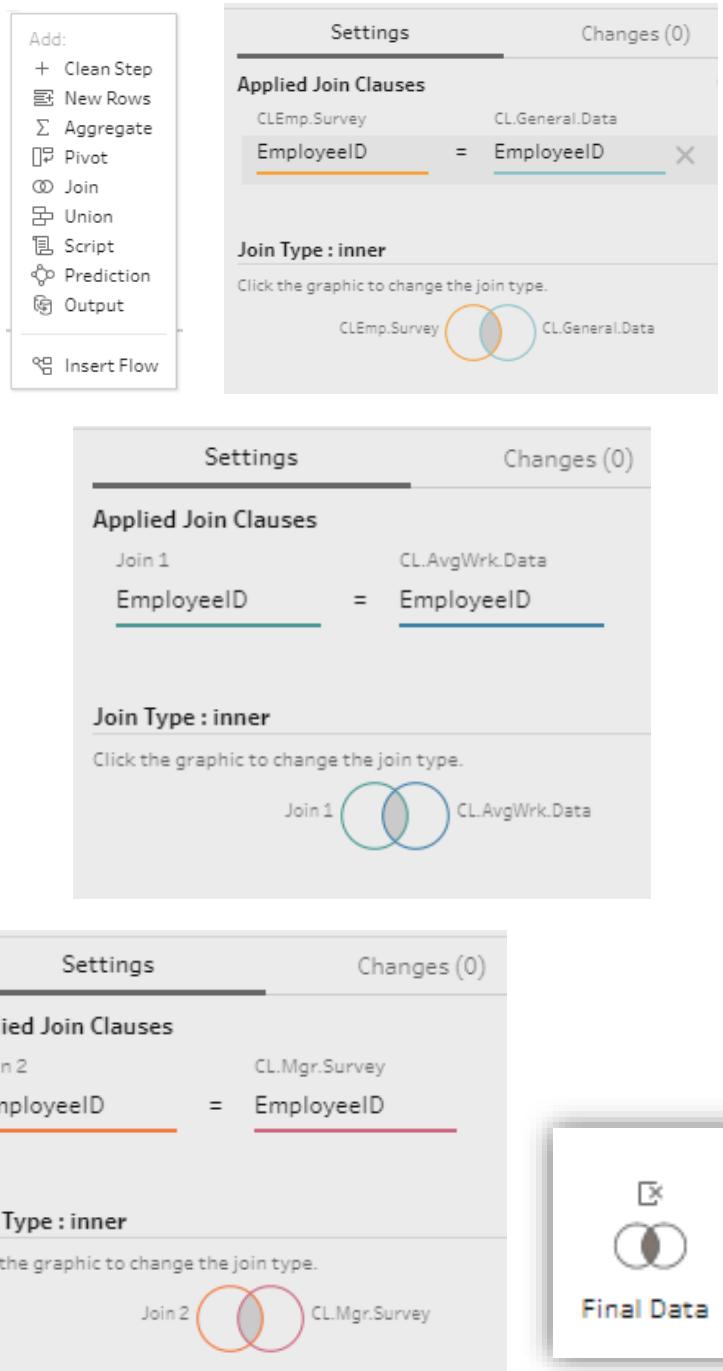


- The first step is to create a calculated field.
- In the popup window an equation is written to convert the seconds to hours.



- Then the previous column “AvgSeconds” is removed from the data.
- Total of 2 changes have been made in this cleaning step.
  The cleaned output is saved in the name “CL.AvgWrk.Survey”.

Joining Multiple Data Files:

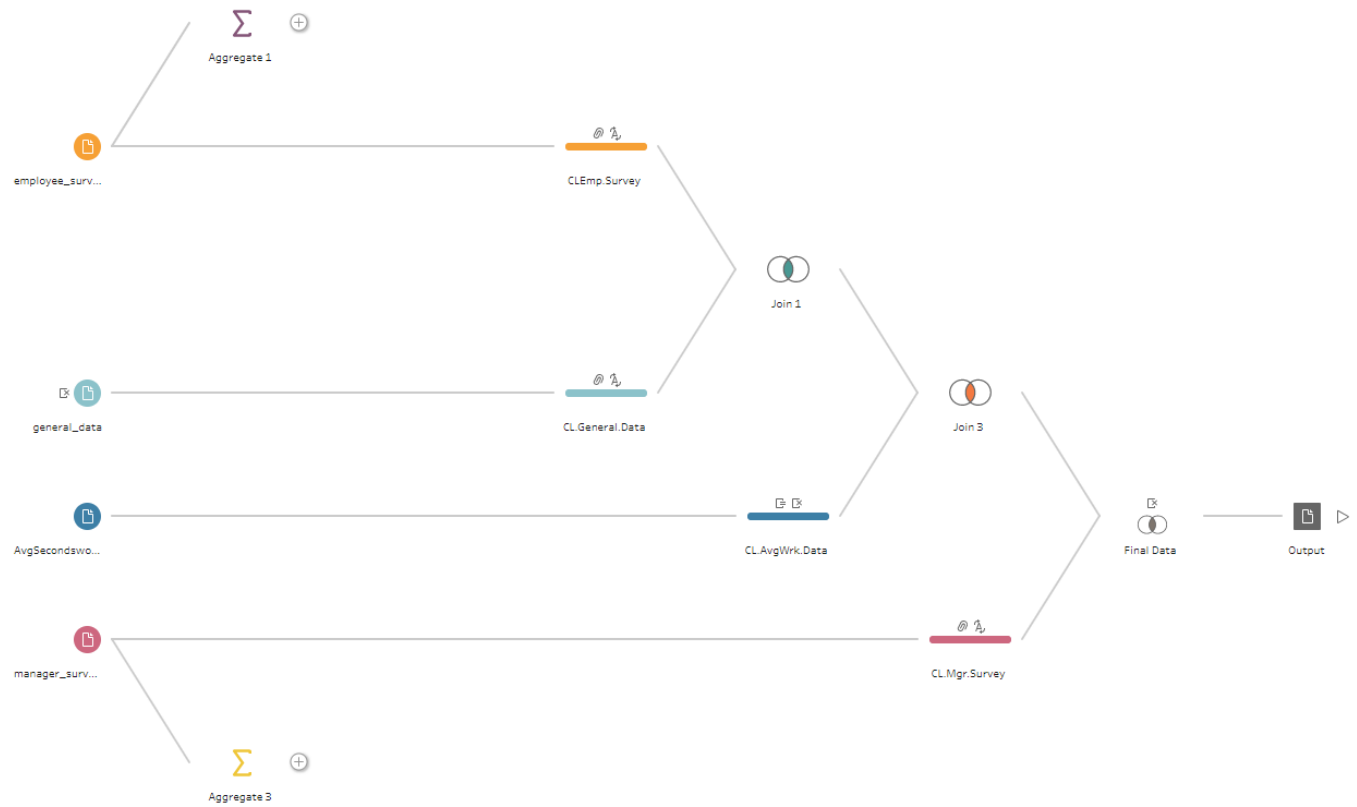


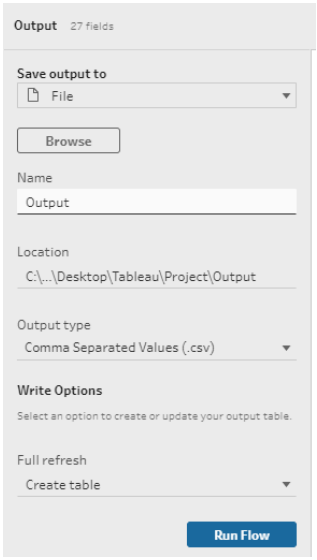
- All the 4 data sets are inner joined on the EmployeeID Column to create the output.
- The cleaned output is saved in the name “Final Data”.

Final Cleaning:

- The Final Join data has 3 different ID columns that has the same value.
- The duplicate ID columns are then selected and removed from the data.
- The cleaned “Final Data” is ready to be saved as csv file.

Saving Data to CSV File:





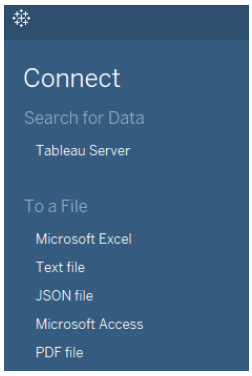
- The cleaned “Final Data” is ready to be saved as csv file.
- The desired file name and file type is chosen.
- “Run Flow” is selected to create the final output csv file.

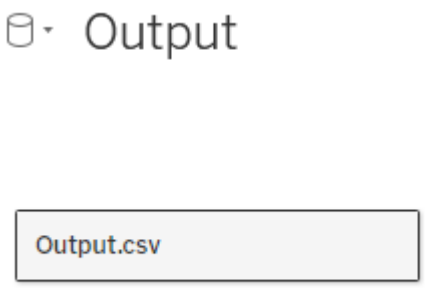
TABLEAU DESKTOP

Tableau Desktop is used for the following:

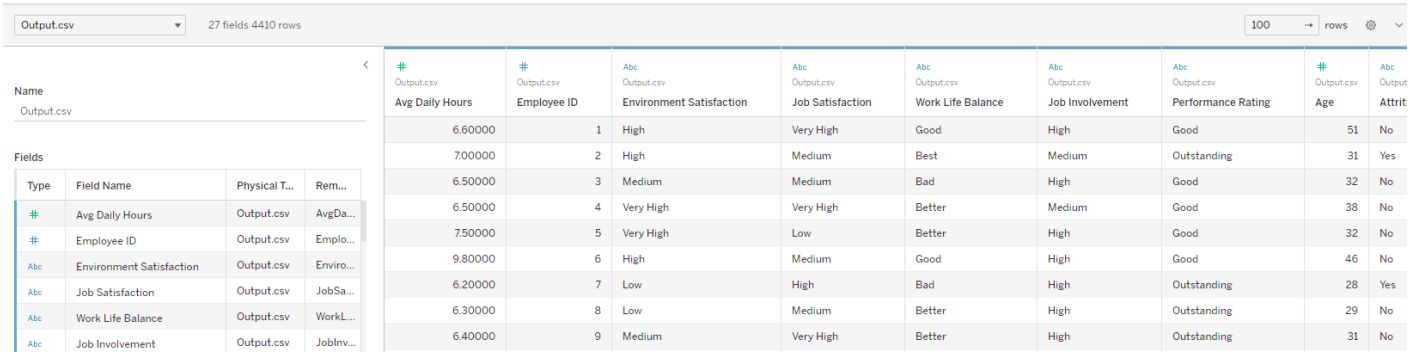
- Creating charts from observational data.
- Creating dashboards for consolidated visualization.
- Creating a story to brief the solution of business problem.

Importing Data from CSV File:



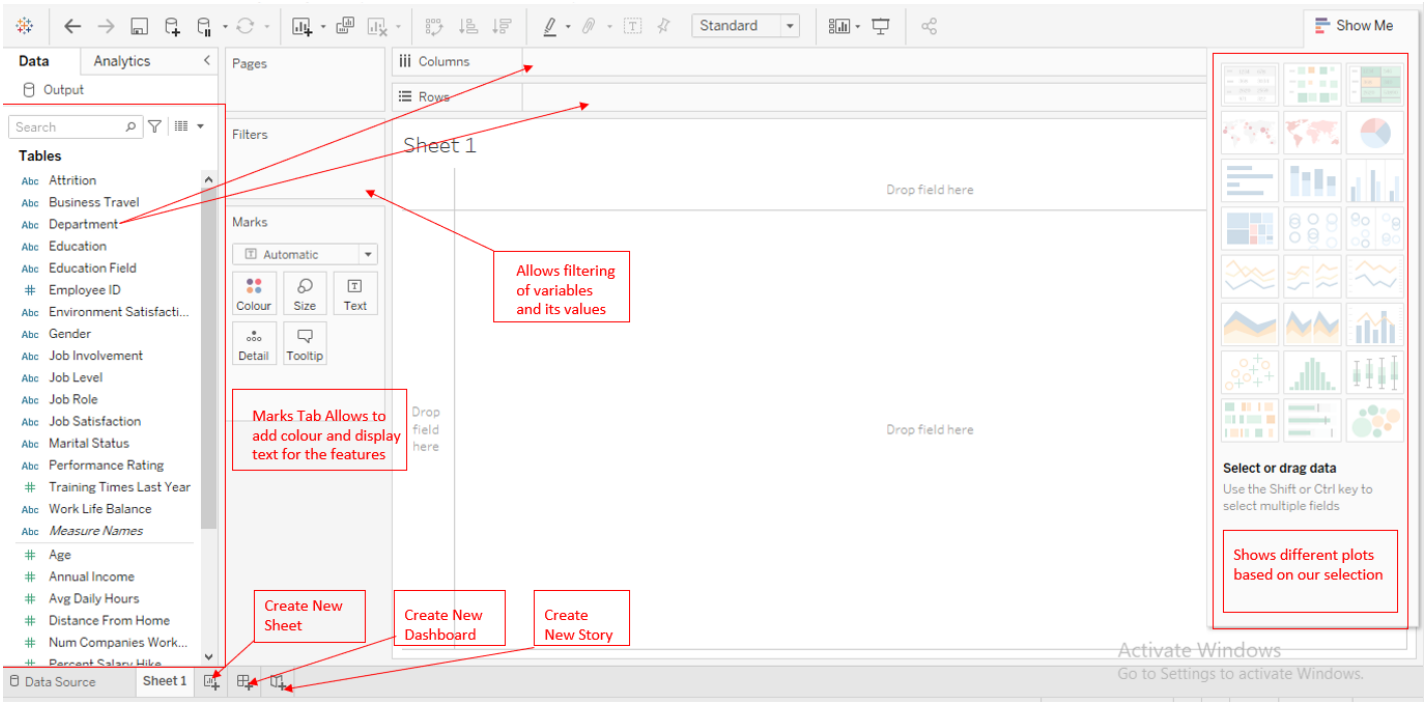


- The data from csv file is imported using text file option in the connection tab.
- Once the data is imported, the fields as well as the data is displayed as in the figure below.



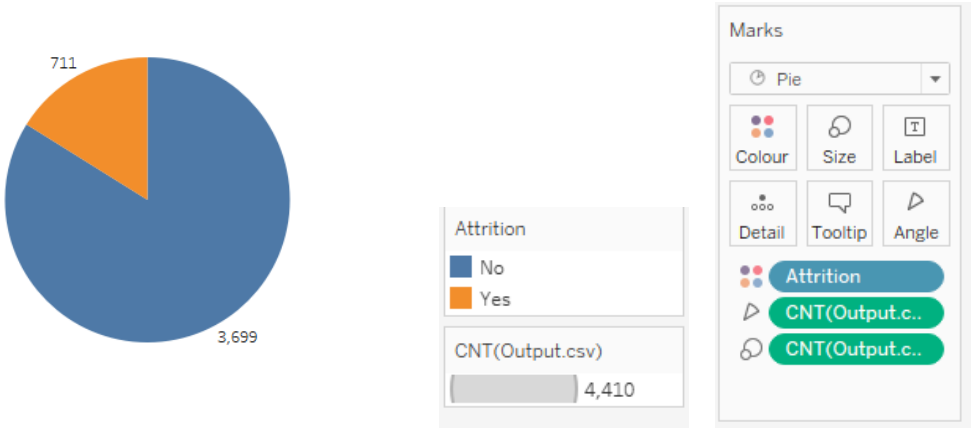
Avg Daily Hours	Employee ID	Environment Satisfaction	Job Satisfaction	Work Life Balance	Job Involvement	Performance Rating	Age	Attrit
6.60000	1	High	Very High	Good	High	Good	51	No
7.00000	2	High	Medium	Best	Medium	Outstanding	31	Yes
6.50000	3	Medium	Medium	Bad	High	Good	32	No
6.50000	4	Very High	Very High	Better	Medium	Good	38	No
7.50000	5	Very High	Low	Better	High	Good	32	No
9.80000	6	High	Medium	Good	High	Good	46	No
6.20000	7	Low	High	Bad	High	Outstanding	28	Yes
6.30000	8	Low	Medium	Better	High	Outstanding	29	No
6.40000	9	Medium	Very High	Better	High	Outstanding	31	No

Familiarizing Desktop Interface:



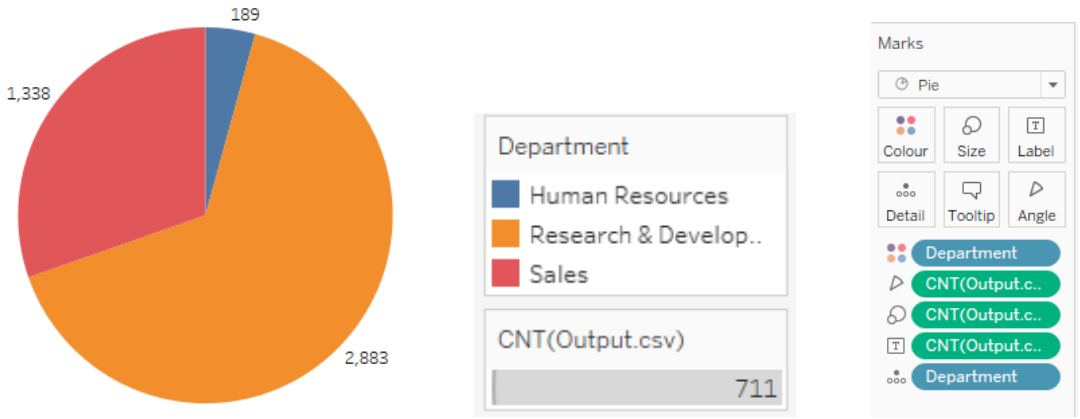
- The features can be dragged and dropped to the rows and columns.
- A filter pane is provided where features could be included so that the plots can be filtered based on the data in that column.
- The different kind of plots are displayed on the right side under Show Me tab where it suggests the kind of plots based on the user feature selection.
- The features are displayed on the left-hand side and are separated based on whether they are categorical or numeric.

Distribution of Attrition:



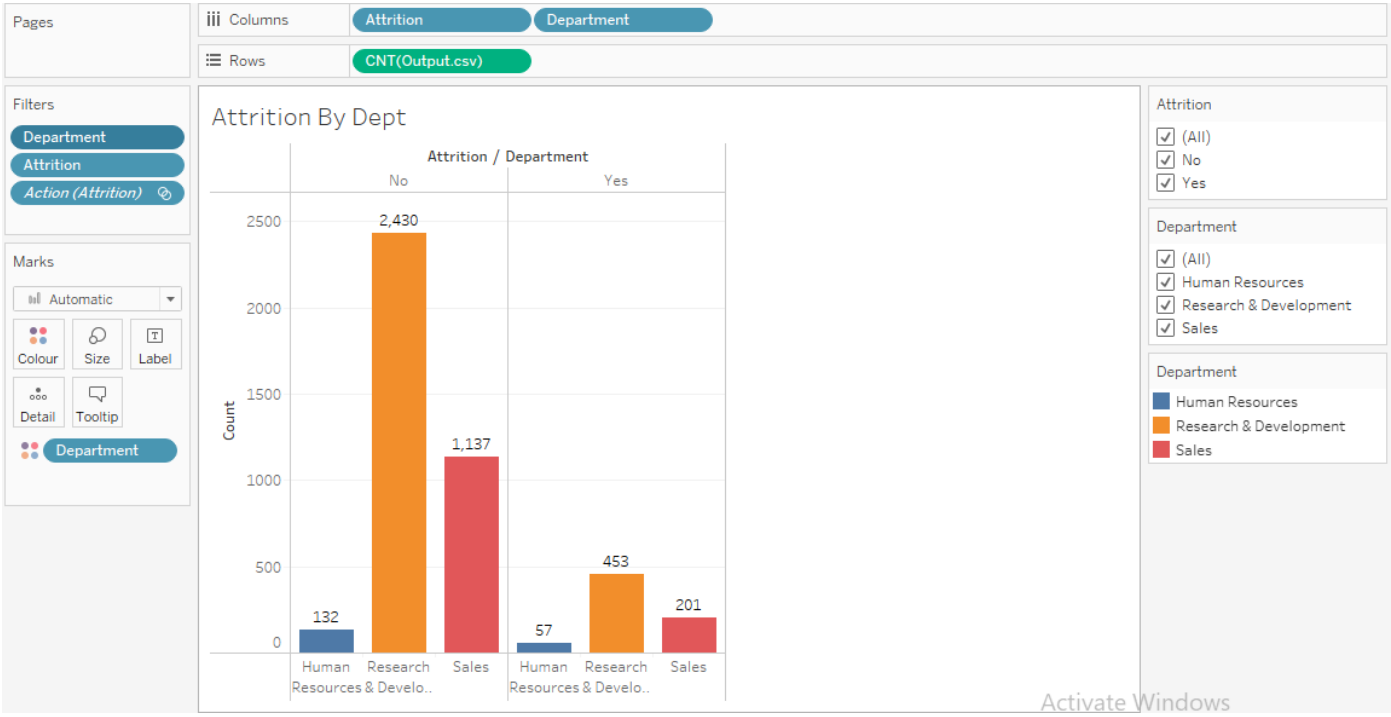
- The Attrition distribution is shown in the pie chart.
- There are 711 employees with attrition “Yes”.
- There are 3699 employees with attrition “No”.
- The total count of employees is 4410.

Distribution of Department:

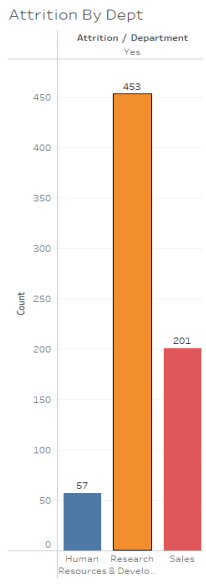


- The departments distribution is as shown in the pie chart.
- The maximum count of employees is in the Research and Development department with count of 2883, followed by Sales [1338] and finally HR [189].

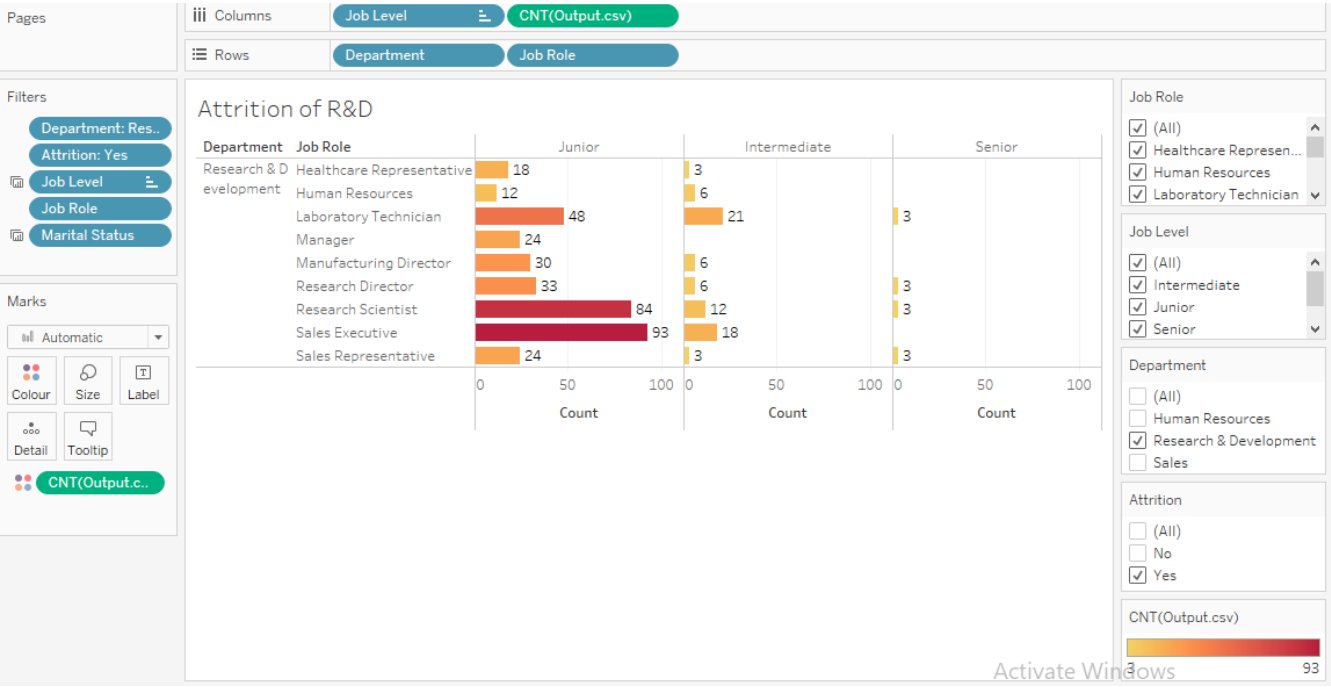
Attrition by Department:



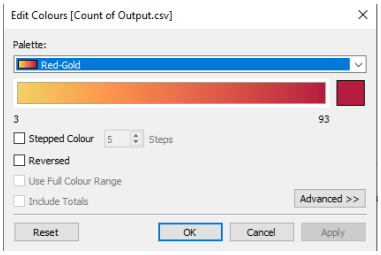
- The department and Attrition features are selected to the columns
- The count is selected to the Rows.
- Filter is employed for Department and Attrition.
- In Marks pane, Department is selected to the colours tab so that each department has different colour.
- The Filters are displayed on the right side and the values can be changed.

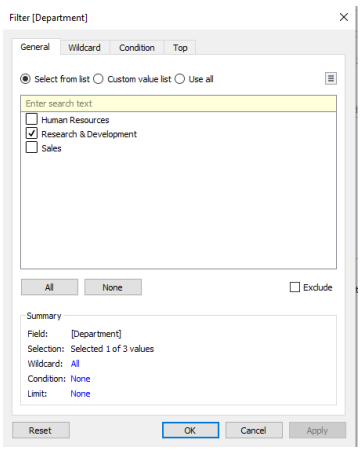


Attrition of Research and Development Department:

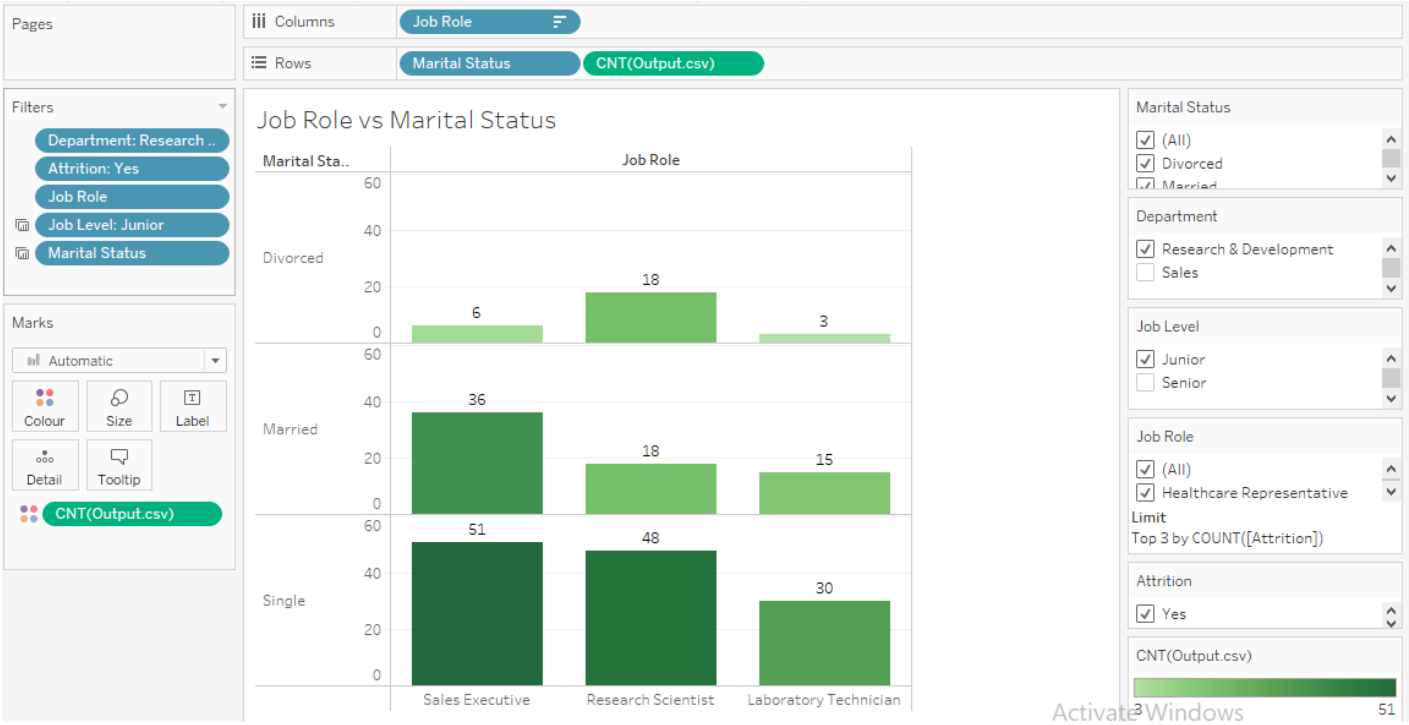


- Job Level and count are selected to the Columns
- Department and Job Role are dragged to the Rows.
- Filters are created for Department, Attrition, Job Level and Marital Status.
- Any combination of the filter values can be tried out for useful insights.
- A gradient coloration is given for the count so that highest values show red colour.

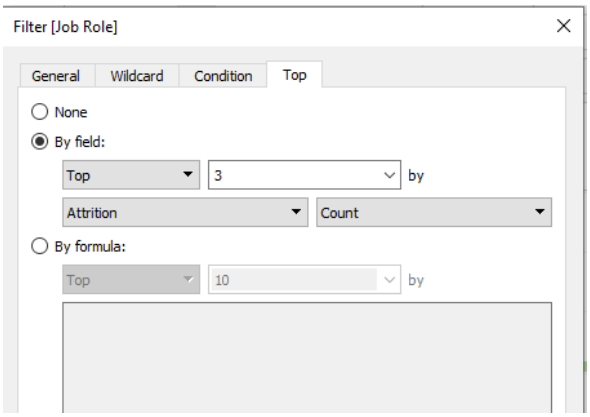




Attrition by Role and Marital Status:



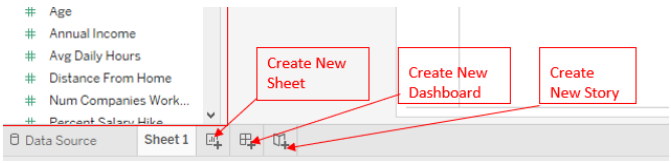
- Job Role is dragged to the Columns.
- Marital Status and count are taken to the Rows.
- Filters are applied for the Department, Attrition, Job Role, Job Level and Marital Status.
- We could see that the maximum attrition is happening for the employees who are single.
- we could see the highest attrition occurred in “Junior” Level and in the job roles : “Laboratory Technician”, “Research Scientist” and “Sales Executive”.
- A gradient coloration is given for the count so that highest values show red colour.

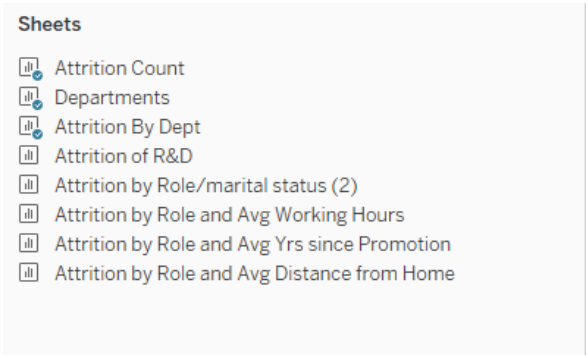


- We have taken the top 3 job roles for by attrition count from the edit filter option as shown above.

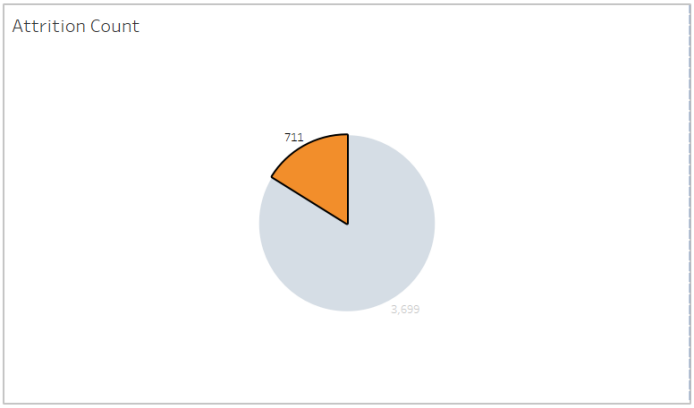
Attrition Dashboard 1 :

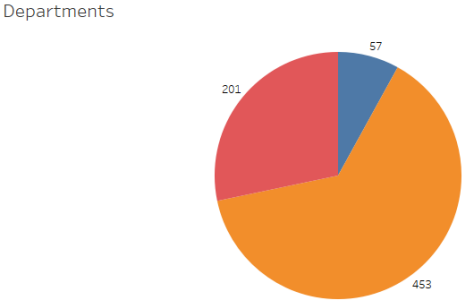
- The New Dashboard is created using the icon as shown below.

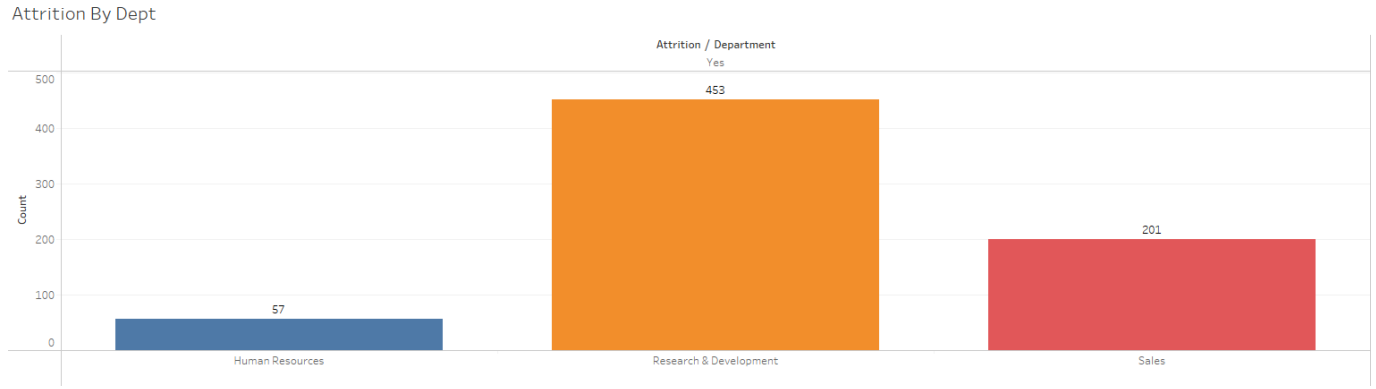




- Drag and drop the required 3 sheets from the list of sheets on left side.





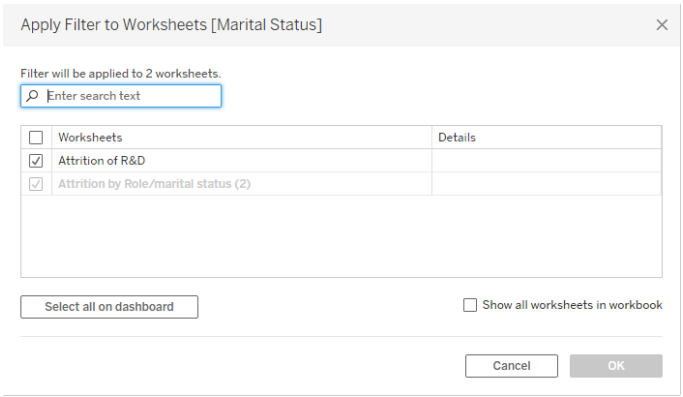


- The plots are interconnected so that when one data change, corresponding changes are reflected in the other.

Attrition Dashboard 2 :

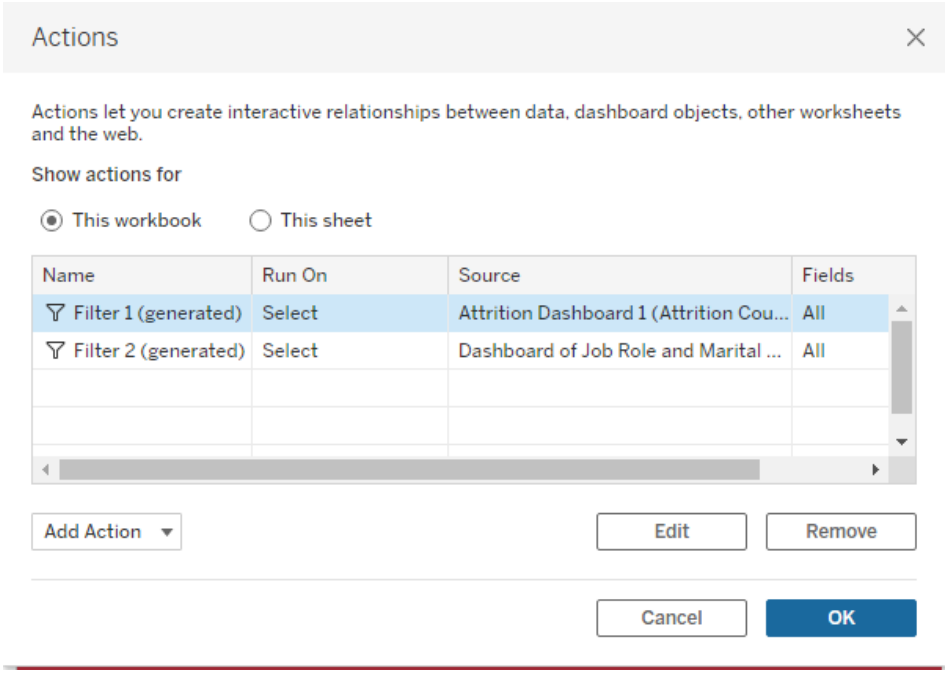
- The Dashboard is created by dragging and dropping 2 sheets to the desired layout positions.

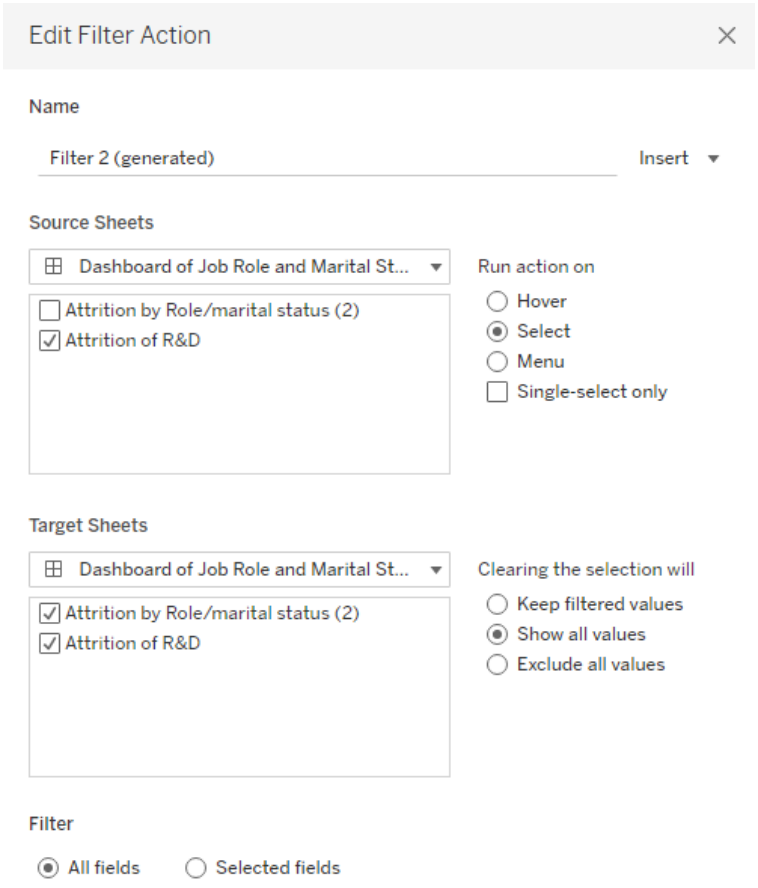
- we could see the highest attrition occurred in “Junior” Level and in the job roles : “Laboratory Technician”, “Research Scientist” and “Sales Executive”.
- The employees who are single exhibited higher attrition than the divorced and married.
- Floating filters of Job level and Marital status is also provided.



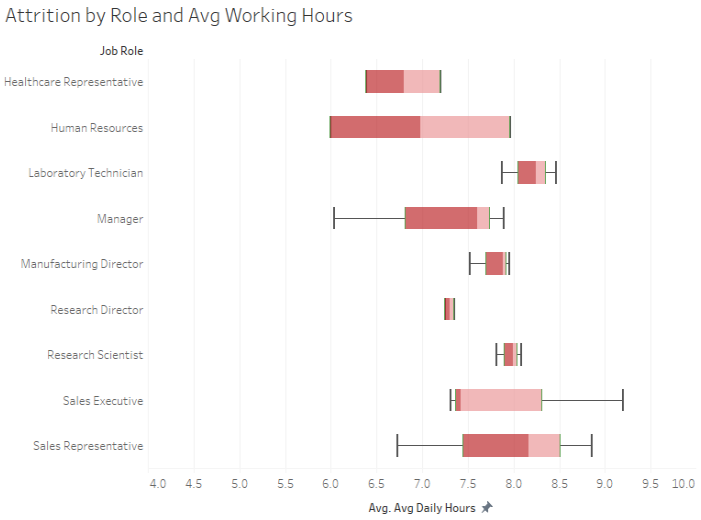
- Above setting is done so as the required sheets get affected by the filter values.

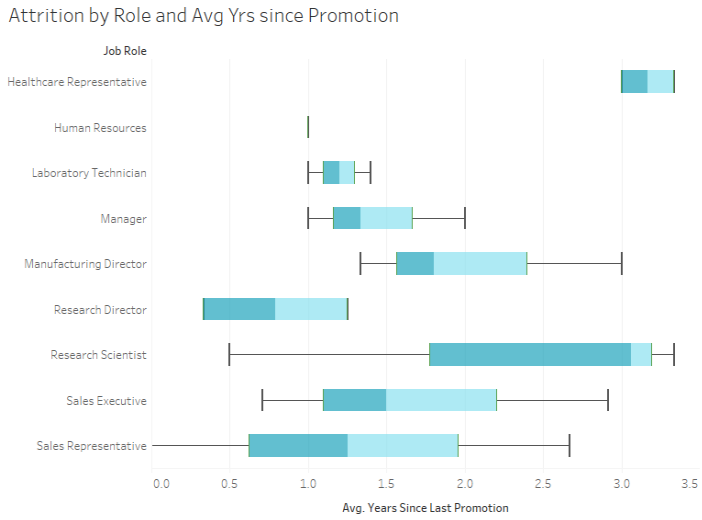
- For the sheets to interact within the dashboard, go to Dashboard -> Actions and edit the filter and change the configuration as required.

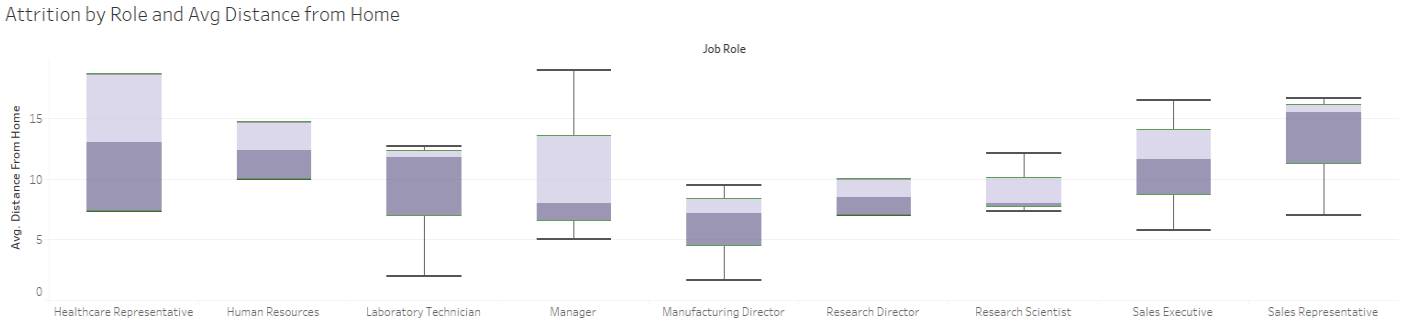




Dashboard 3: Reasons for Attrition

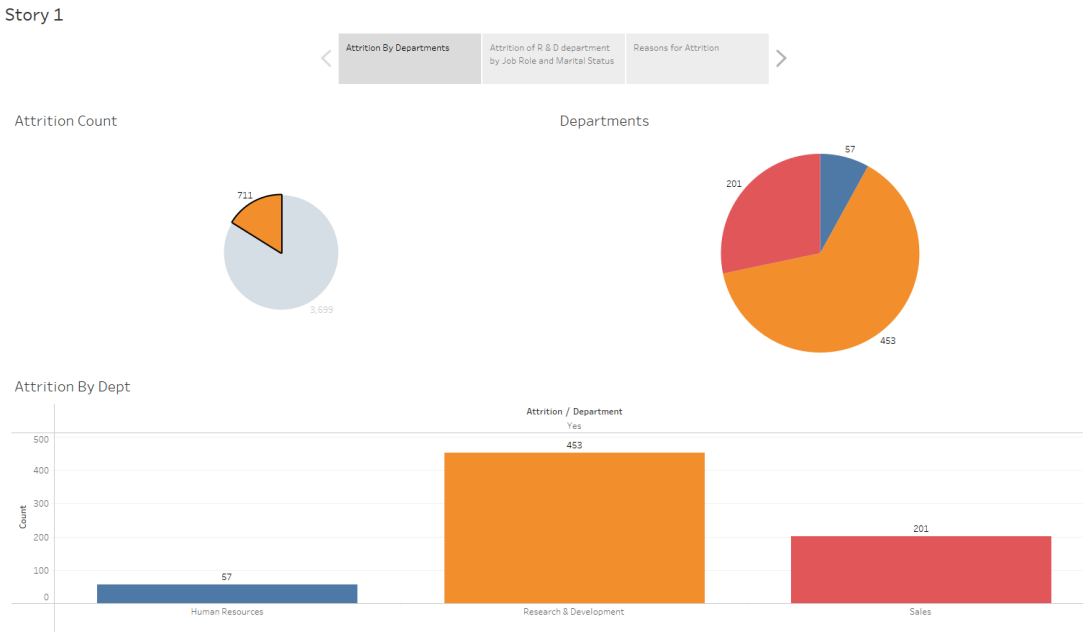




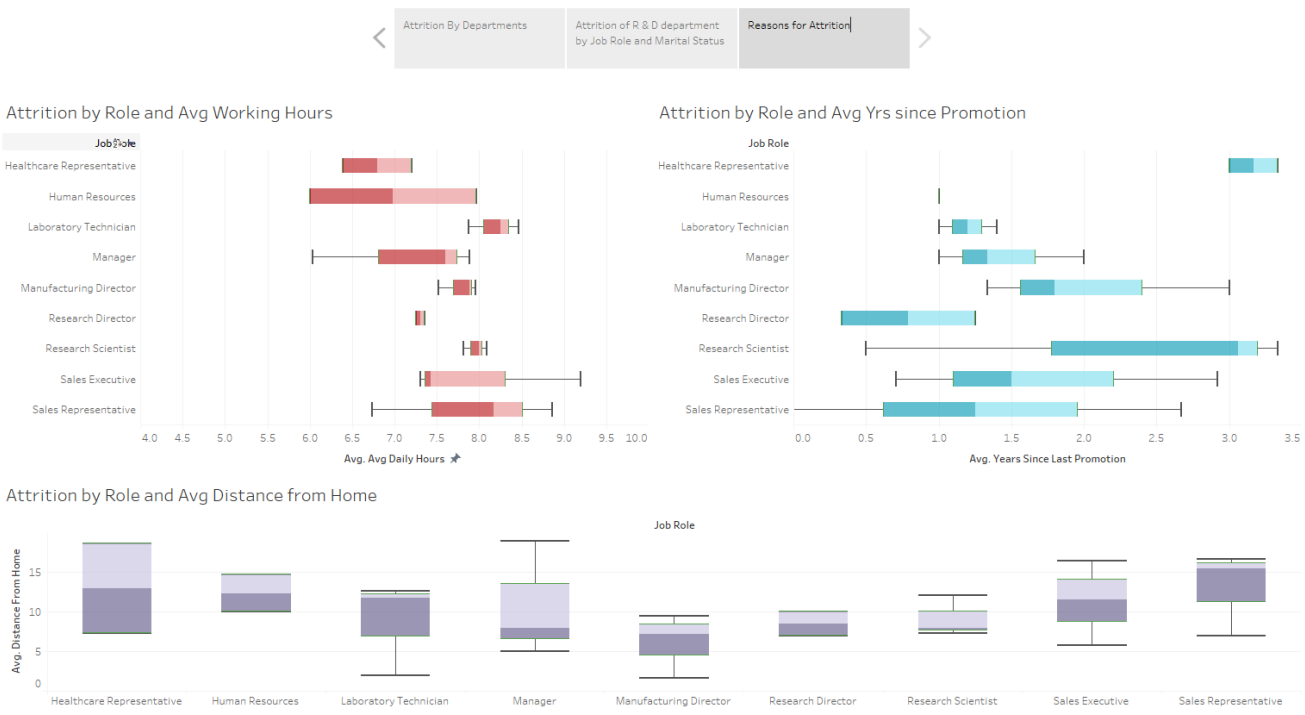


- Box plots are created to analyze the possible reasons for attrition.
- We could see that the Average working hours, Years since promotion and Distance from home are the main factors that affected the employee attrition.

Story of Attrition:



Story 1



- From the department distribution -> Job role and Marital Status -> Major Reasons

CONCLUSION

From the analysis we could conclude the following:

- Attrition is highest in the Research & Development Department.
- In the Research & Development department, the junior employees and employees who are single are more subjected to attrition.
- The top 3 Job roles that had highest attrition is “Laboratory Technician”, “Research Scientist” and “Sales Executive”.
- The major reasons for attrition among the above category are high Average working hours, high Years since promotion and long Distance from home.

RECOMMENDATIONS

- Reduce the working hours of the employees or increase the overtime pay if they need to work longer.
- Increase the promotion frequency or make the employees happy with some bonus/profit sharing for their hard work.
- Entertain car pooling system or provide travel allowance for employees staying far as above 10km from the office.