

In this project I worked with a sophisticated real-life database, that is an Employee database. It consisted of 5 different tables named employees, department, dept manager, dept_emp, salaries with information about different employees, managers, their departments, salaries and working periods. The main objective of this project was to analyze 'employees' database using Mysql and integrating it with tableau to Visualize and find patterns between male-female salaries across the company. I have analyzed the data using Select, Insert, Update, and Delete statements, aggregate functions, joins, views, and other sub-queries. I performed various sql queries to get different insights from the data. To visualize and analyze the Solution in Tableau, Dataset was transferred from SQL and thus integrated SQL-Tableau. Did thorough department-wise analysis of male and female employees, the average salary of female versus male employees in the entire company. I have organized different charts into beautiful Interactive Dashboards and developed these insights:

1. Avg salary of both male and female employees is increasing over the period. But the pay gap is also increasing btw male and females. Females are paid less than the males.
2. The increase in no. of employees is significant at the initial periods but the number of employees are the same at the later stage showing no increase in no. of employees. This might be due to no new hirings or may be due to no. of employees leaving the company is the same as joining.
3. The ratio of no. of male and female employees was nearly in the 2:3 ratio.
4. There are more no. of male managers than female managers.
5. Avg salary order: sales>finance>marketing, least for human resources dept.
6. After what year, the salary gap is increasing?

After performing EDA on employee data, I written sql queries to gather the data to transfer to tableau. I integrated tableau and MySQL workbench through an inbuilt feature offered by tableau. Then I established the star schema which is a technique of establishing relationships between the table so that tableau can perform its in built calculations.

And finally compiled them all to make interactive dashboards showing breakdown between male and females, number of managers, average salary.

Visualizations:

- What is the breakdown between the male and female employees working in the company each year, starting from 1990? (stacked bar chart)
- What is the number of male managers to the number of female managers from different departments for each year, starting from 1990? (line chart)
- What is the average salary of female versus male employees in the entire company until year 2002? Need to filter out by departments. (Stacked line chart)
- What is the average male and female salary per department within a certain salary range? (Bar charts)

SQL-Tableau Integration Project:

SQL helps us store and manipulate the data we are working with, and **Tableau** enables beautiful data visualization. A well-thought-out integration stepping on these two pillars could give us insights that we have not anticipated.

In this project I worked with a sophisticated real-life database, that is an Employee database. It consisted of 5 different tables, with information about different employees, managers, their departments, salaries and working periods. I have analyzed the data using Select, Insert, Update, and Delete statements, aggregate functions, joins, views, and other sub-queries. To visualize and analyze the Solution in Tableau, Dataset was transferred from SQL and thus integrated SQL-Tableau. I have organized different charts into beautiful Interactive Dashboards and developed these insights:

Insights:

1. Avg salary of both male and female employees is increasing over the period. But the pay gap is also increasing btw male and females. Females are paid less than the males.
2. The increase in no. of employees is significant at the initial periods but the number of employees are the same at the later stage showing no increase in no.of employees. This might be due to no new hirings or may be due to no.of employees leaving the company is the same as joining.
3. The ratio of no.of male and female employees
4. was nearly in the 2:3 ratio.
5. There are more no. of male managers than female managers.
6. Avg salary order: sales>finance>marketing, least for human resources dept.
7. After what year, the salary gap is increasing?

Visualizations:

- What is the breakdown between the male and female employees working in the company each year, starting from 1990?(stacked bar chart)
- What is the number of male managers to the number of female managers from different departments for each year, starting from 1990?(line chart)
- What is the average salary of female versus male employees in the entire company until year 2002? Need to filter out by departments.(Stacked line chart)
- What is the average male and female salary per department within a certain salary range?(bar charts)

What impacted you to work on this project? Having worked with ML and data modeling in my previous project, I wanted to learn more about how to present data to make it more interpretable and visually appealing and how to visualize it in a better way.

1. What is data analysis?

Data analysis is the science of examining a set of data to draw conclusions about the information to be able to make decisions or simply to expand the knowledge on various subjects. It consists of subjecting data to operations. Just like we did in this project. We analysed the given data and digged some relevant insights from it.

2. What do you mean by Tableau? Where do we use it?

Tableau is a powerful and fastest growing data visualization tool used in the Business Intelligence Industry. It helps in simplifying raw data in a very easily understandable format. Tableau helps create the data that can be understood by professionals at any level in an organization.

3. Why did you integrate SQL and Tableau?

Using SQL and Tableau together takes the data analysis to the next level. We can easily connect the SQL Server to Tableau and extract the data directly into it. To visualize data better using Tableau as SQL is the base database.

4. What were the different aggregate funcs used?

- a) AVG-calculates average of a set of values
- b) COUNT-counts row in a specified table/view
- c) SUM- sum of the values
- d) MIN-minimum value in a set of values
- e) MAX-maximum value in a set of values

5. What are joins? Name the types (Study self-join)

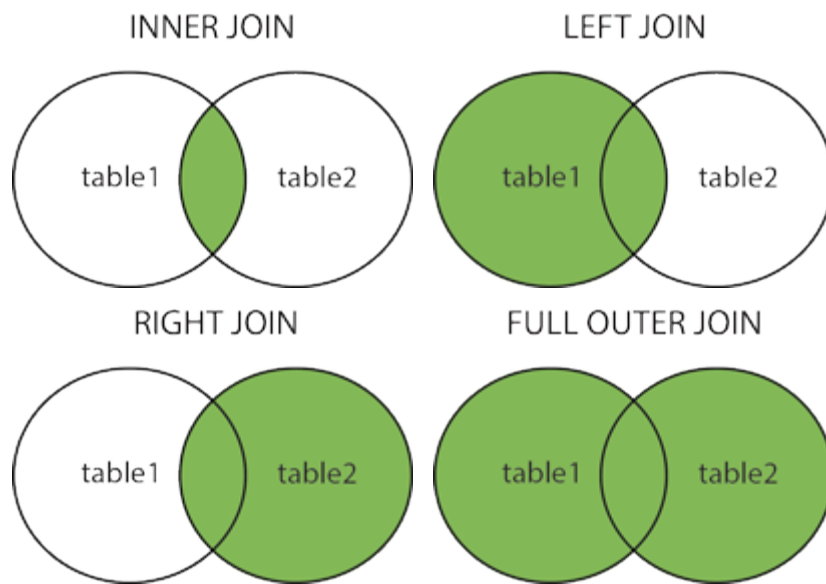
3 different types of joins:

- a) Inner join: return records having maximum values in both the tables

Right join: return all the records from the right table and matching records from the left table

- b) Left join: return all the records from the left table and matching records from the right table

- **(INNER) JOIN**: Returns records that have matching values in both tables
- **LEFT (OUTER) JOIN**: Returns all records from the left table, and the matched records from the right table
- **RIGHT (OUTER) JOIN**: Returns all records from the right table, and the matched records from the left table
- **FULL (OUTER) JOIN**: Returns all records when there is a match in either left or right table(union)



6. What do you mean by important metrics? How were they identified?

A metric is a quantifiable measure that is used to track and assess an aspect of business performance. Metrics are utilized to track the progress made towards both short and long-term goals.

The most important metrics that a business tracks are called key metrics. Also known as a key performance indicator, or KPI, a key metric is a statistic which, by its value gives a measure of an organization or department's overall health and performance. KPIs, or key performance indicators, are vital metrics connected to time sensitive goals. All organizations, from nonprofits to multinational corporations, need to track key metrics in order to have an accurate picture of what is going on.

Important parameters used in the project were employees, managers, their departments, salaries and working periods.

Metrics: gender ratio in the senior segment position

Employee count per dept

7. What do you mean by interactive dashboards? How do we use it and how are they presented?

An interactive dashboard allows you to drill down and filter operational information so data can be viewed from different perspectives or in more detail. With a condensed and clear presentation of the overall business figures, dashboards enable data-driven company decisions. With interactive dashboards that enable you to visualize your data, filter on demand and simply click to dig deeper into the underlying data.

Easy to use and easy to infer !!

8. What was the nature of the employee/sales database?

5 tables, 300,000 employees, salaries, depts., duration of working.

Features?
length of the dataset?

9. How did you visualize and analyze the male/female salaries?
average salaries, agg function and group by clause.

10. What were the major queries you have used from SQL and which were the most useful ones?

The major queries I have used are joins, as there were different tables and I had to take insights from the combination of these tables. The other queries are aggregate functions of Average and Sum, these I used to get the average salaries and total employees present according to gender.

11. No of charts and name the axes labels of the charts (insights from the charts)
12. Select, Insert, Update, Delete and aggregate functions.

Update- specify the name of the table, and change data using update clause.

Insert- insert into table values

Delete The DELETE statement is used when we want to remove some or all of the records from the table, while the TRUNCATE statement will delete all the rows from a table since we can't use where command with truncate. DELETE is a DML command as it only modifies the table data, whereas the TRUNCATE is a DDL command.

Drop is used to drop the entire structure of the table. Table loses its entire existence.

Drop is also a DDL command.

Aggregate function used with group by clause.

13. What are views (Tableau+SQL) and what are subqueries?

A view is a visualization or viz that you create in Tableau. A viz might be a chart, a graph, a map, a plot, or even a text table.

Subqueries –inner query nested query with another sql query and embedded within the where clause.

14. How did you transfer the database from SQL? How did you integrate it with Tableau?
Sql tables were converted to csv and then integrated to tableau. Tableau's in-built SQL Server connector conveniently establishes a connection with SQL Server. You can easily set up a data source and use Tableau to visualize your data to perform a fruitful analysis for your business. Inbuilt function in tableau. Code in tableau.

15. Insights on dept wise analysis.

There are 9 departments in total. They are: Customer service, Development, Finance, Human resources, Marketing, Production, Quality and Management, Research, Sales. Avg salary order: sales>finance>marketing, least for human resources dept. Avg salary of male and female employees is same over the diff departments. Though there are differences in avg salary in different departments.

16. What is SQL, MySQL, MySQL workbench, Tableau? Where do we use it?

MySQL is an open source relational database that is cross platform. MySQL workbench is an integrated development environment for MySQL server. It has utilities for database modeling and designing, SQL development and server

administration. MySQL Workbench is a unified visual tool for database architects, developers, and DBAs. MySQL Workbench provides data modeling, SQL development, and comprehensive administration tools for server configuration, user administration, backup, and much more.

17. Number of databases used? Which were the different database tools used? Different visualization tools used in the project

1 database.(Employees Mod) Tableau, Mysql.

18. Alternatives of Tableau? Alternatives of MySQL workbench?

Alternative of tableau- power BI,

Alternative of mysql workbench- HeidiSQL

19. Tableau vs Power BI

Tableau platform is known for its data visualization functionality whereas Power BI offers numerous data points to offer data visualization. Tableau BI can handle a huge volume of data with better performance while Power BI can **handle a limited volume of data**.

20. What is DDL, DML, DCL?

DDL: Data definition language –truncate drop create alter

DML: Data manipulation language- delete,insert,update,select

DCL: Data control Language-grant, revoke(uses of them)

TCL: Transaction Control Language- Commit, Rollback, start transaction

21. What is a relational database?

A relational database is a type of database that stores and provides access to data points that are related to one another. ... The columns of the table hold attributes of the data, and each record usually has a value for each attribute, making it easy to establish the relationships among data points. SQL is a programming language that is used by most relational database management systems (RDBMS) to manage data stored in tabular form (i.e. tables).

22. **Tech interview for SQL and Tableau(Top 100 Tableau and SQL interview questions)**, primary key, foreign key, different operators(AND, OR...), wildcard characters, order by, group by, having, while, where, limit

23. Skills gained from this project (Tech + Business understanding + **how to present data**)

24. Real-time applications of this project

Employees are present in all the companies. Employee growth plays a crucial role in the growth and prosperity of the company. Mostly the companies will be working on client projects or any services but this project can be used to analyse the employees within the company and can get more insights regarding their employees like the

male, female categorization, new hirings, their salaries. The increment in all these clearly depicts the growth of the company

25. Challenges faced during this project

not aware of the feature of integration (developed curiosity) had to learn about tableau from start

26. What are dashboards? What are stories?

A dashboard is **a collection of several views, letting you compare a variety of data simultaneously**. For example, if you have a set of views that you review every day, you can create a dashboard that displays all the views at once, rather than navigate to separate worksheets.

Use stories to make your case more compelling by showing how facts are connected, and how decisions relate to outcomes. You can then publish your story to the web, or present it to an audience.

Each story point can be based on a different view or dashboard, or the entire story can be based on the same visualization seen at different stages, with different filters and annotations

In **Tableau**, a **story** is a sequence of visualizations that work together to convey information. You can create **stories** to tell a data narrative

27. Wild Characters in SQL: A wildcard character is used to substitute one or more characters in a string. Wildcard characters are used with the [LIKE](#) operator. The **LIKE** operator is used in a **WHERE** clause to search for a specified pattern in a column.

28. **Difference between Primary Key and Foreign Key:**
The purpose of the foreign key is to ensure referential integrity of data.

PRIMARY KEY

A primary key is used to ensure data in the specific column is unique.

FOREIGN KEY

A foreign key is a column or group of columns in a relational database table that provides a link between data in two tables.

- | | | |
|---|---|--|
| 2 | It uniquely identifies a record in the relational database table. | It refers to the field in a table which is the primary key of another table. |
| 3 | Only one primary key is allowed in a table. | Whereas more than one foreign key are allowed in a table. |
| 4 | It is a combination of UNIQUE and Not Null constraints. | It can contain duplicate values and a table in a relational database. |

A foreign key is a column (or columns) that references a column (most often the primary key) of another table. The purpose of the foreign key is to ensure referential integrity of the data. In other words, only values that are supposed to appear in the database are permitted.

For example, say we have two tables, a **CUSTOMER** table that includes all customer data, and an **ORDERS** table that includes all customer orders. Business logic requires that all orders must be associated with a customer that is already in the **CUSTOMER** table. To enforce this logic, we place a foreign key on the **ORDERS** table and have it reference the primary key of the **CUSTOMER** table. This way, we can ensure that all orders in the **ORDERS** table are related to a customer in the **CUSTOMER** table. In other words, the **ORDERS** table cannot contain information on a customer that is not in the **CUSTOMER** table.

The structure of these two tables will be as follows:

Table **CUSTOMER**

Column Name	Characteristic
SID	Primary Key
Last_Name	
First_Name	

Table **ORDERS**

Column Name	Characteristic
Order_ID	Primary Key
Order_Date	
Customer_SID	Foreign Key
Amount	

In the above example, the Customer_SID column in the **ORDERS** table is a foreign key pointing to the SID column in the **CUSTOMER** table. The following examples are operations that violate the referential integrity of this relationship:

Inserting a row in the **ORDERS** table where Customer_SID does not appear in the SID column in the **CUSTOMER** table.

Deleting a row from the **CUSTOMER** table where the SID of the row to be delete is still present in the Customer_SID column in the **ORDERS** table.

When these operations are attempted, the database would return an error stating that referential integrity is violated.

It is worth noting that foreign keys allow NULL, while the primary key does not. In addition, a foreign key does not always have to reference a primary key of another table. It can also reference a column that has the UNIQUE constraint.

29. Business recommendations you thought of after developing insights.

Employees are present in all the companies. Employee growth plays a crucial role in the growth and prosperity of the company. Mostly the companies will be working on client projects or any services but this project can be used to analyze the employees within the company and can get more insights regarding their employees like the male, female categorization, new hirings, their salaries. The increment in all these clearly depicts the growth of the company.

30. **Group By , Order By, Having:**

```
SELECT SUBJECT, YEAR, Count(*)
FROM Student
GROUP BY SUBJECT, YEAR;
```

(Select->agg function->group by-> having)

SELECT NAME, SUM(SALARY) FROM Employee

GROUP BY NAME

HAVING SUM(SALARY)>3000;

31. Difference between Having and where:

A HAVING clause is like a WHERE clause, but **applies only to groups as a whole** (that is, to the rows in the result set representing groups), whereas the WHERE clause applies to individual rows. ... You can apply a HAVING clause only to columns that also appear in the GROUP BY clause or in an aggregate function.

32. What is the difference between coalesce () and Isnull ()?

With COALESCE, the resulting column is defined as NOT NULL only if all expressions are non nullable and NULL otherwise. With ISNULL, the resulting column is defined as NOT NULL if any expression is non nullable and NULL if all expressions are nullable.

33) Alias- to give a temporary name to the object.