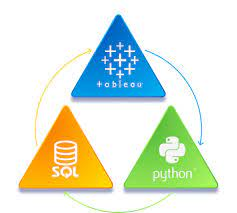
**Salary Analysis with Pandas, SQL and Tableau**



# **General İnformation**

It is crucial to check the analysis results with different tools. If we aren’t sure about the analysis result with one tool, we can check the result with another data science tool. Python and SQL are essential tools for handling data properly for this purpose. Hence, we should know more than one data science tool.

In this article, we are going to analyze the salary file with both Pandas and SQL and Tableau, respectively, for the same question. There are lots of SQL tools, like [MS SQL Server](https://learn.microsoft.com/en-us/sql/sql-server/?view=sql-server-ver16), [MySQL](https://www.mysql.com), [PostgreSQL](https://www.postgresql.org), [MongoDB](https://www.mongodb.com), [Oracle](https://www.oracle.com), [SQLite](https://www.sqlite.org/index.html), [Cassandra](https://cassandra.apache.org/_/index.html), [Neo4j](https://neo4j.com), [CouchDB](https://couchdb.apache.org), [MariaDB](https://mariadb.org), [DB2](https://www.ibm.com/products/db2), and [Redis](https://redis.com). In this article, I used MS SQL Server. I suppose that I loaded the “Salaries.csv” file into the MS SQL Server database and Tableau. If you wonder how to read the file from a SQL Server database via Python code and open the file in Tableau from a SQL Server database, you may read my related article [here](https://medium.com/@mevltyldz1/analyzing-data-with-python-sql-and-tableau-simultaneously-ab3154945eaa). The related file can be found [**here**](https://github.com/MevlutYildiz77/MyOwnProjects/blob/master/Salaries.csv) in the project's GitHub repository.

But let's start by learning the fundamentals of these programming languages.

Exactly what is SQL?

Data scientists may manage databases, collect data from them, and create their own databases using the computer language known as SQL, or Structured Query Language. Many businesses use relational databases to store their data because they can connect data from one table to another by using columns, rows, and tables. In order to build and maintain these databases, SQL is the most commonly used language. Data scientists also use SQL to rapidly analyze data, produce data insights, and extract records from massive databases. Databases provide several advantages over apps, websites, and business software programs.

What is Python, exactly?

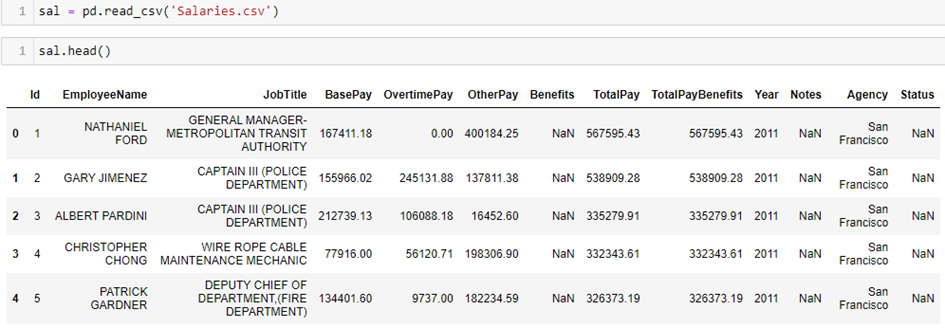
Python is a multipurpose programming language that can be applied to a wide range of tasks, including system scripting, back-end development, and more. Python is frequently used by data scientists to create data analysis tools because of its straightforward syntax and appeal to businesses. Because it can be used on several platforms and places a strong focus on readability, Python is one of the most widely used languages for data exploration.

What is Tableau, exactly?

Today's industry makes extensive use of the data analytics and visualization platform Tableau. Many businesses consider it essential for jobs involving data science. Tableau's drag-and-drop interface contributes to its user-friendliness. This function makes it incredibly simple and quick to complete activities like sorting, comparing, and analyzing. Tableau is a great option for data scientists since it works with a variety of sources, including Excel, SQL Server, PDF files, spatial files, JSON files, statistical files, and cloud-based data repositories.

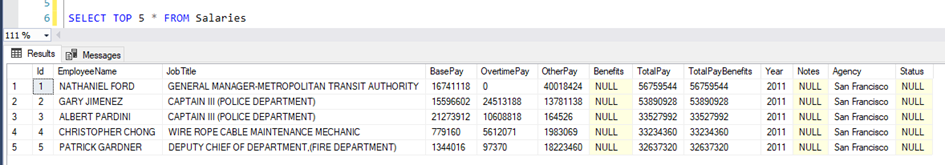
# **2. Reading the File**

First, let us read the “Salaries.csv” file for the first 5 rows with both Pandas and SQL. To open the file with Pandas, I first use the “[pd.read\_csv](https://pandas.pydata.org/docs/reference/api/pandas.read_csv.html" \l "pandas.read_csv)” function, and to see the first five rows, I use the “[head()](https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.head.html" \l "pandas.DataFrame.head)” function.

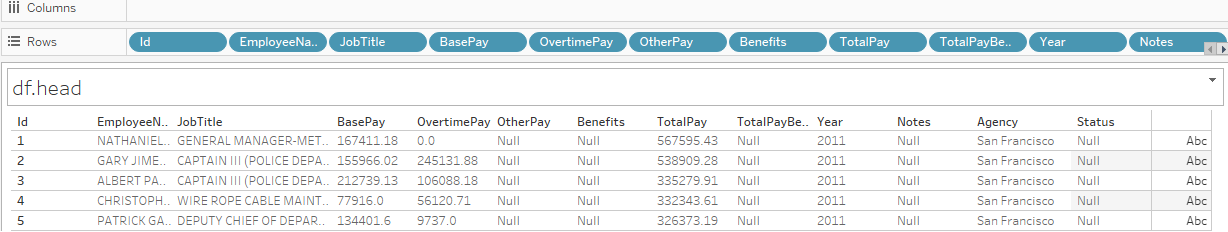
****

**Table 1: Read File With Pandas**

To read the file with SQL, I use the “SELECT \* FROM Salaries” query, and to see the first five rows, I use the “TOP 5” statement.

****

**Table 2: Read File with MS SQL Server**



**Table 3: Read File with Tableau**

This table contains information about the employees' salaries according to their Job Title, Total Pay (total of BasePay, OvertimePay, and OtherPay), TotalPayBenefits (total of TotalPay and Benefits), Year, Agency, and Status. This table has 13 columns.

# **3. Beginning the Analysis**

Now, it is time to analyze. We want to show two examples. One is to analyze the *“number of different years and how many records there are in each year, from the biggest to the lowest.”*  The other is *“how many people have the word ‘manager’ in their job title.”*

### Example 1: Number of different years and how many records there are in each year, from the biggest to the lowest.

It may change the each year’s records year by year. So, we wonder how many there are different years in the file and how many records there are in each year. We show the analysis results using Python, SQL, and Tableau, respectively.

## **Step-1: Show With Python**

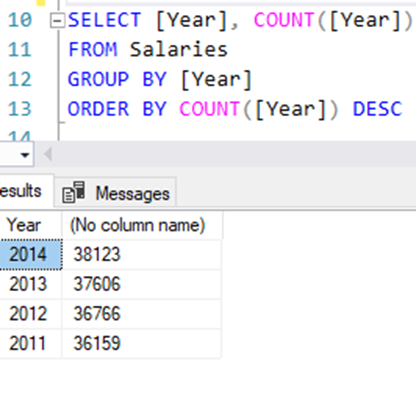
## 

## 

**Table 4:Number of Different Years and Records Number with Python**

## **Step-2: Show with SQL**

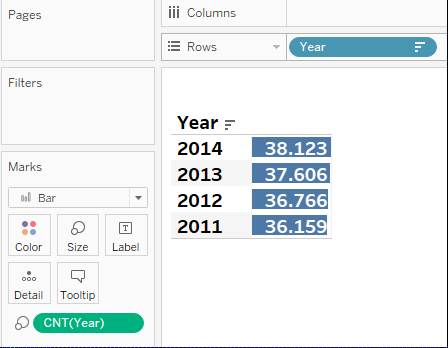
## 

****

**Table 5:Number of Different Years and Records Number with SQL**

#### **Step-3: Show With Tableau**

## 

****

**Table 6:Number of Different Years and Records Number with Tableau**

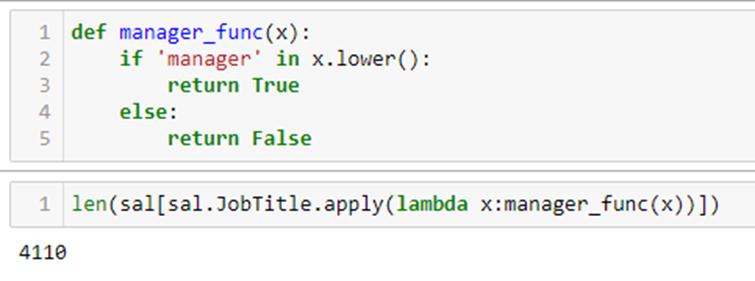
As we see, the analysis results are the same for Pandas, SQL, and Tableau. There are four different years, and their records are the same for each year.

### Example 2: Find how many people have the word “manager” in their job title

We suppose that we want to learn how many people have the word “manager” in their job title. So, we show the results of analyzing Python, SQL, and Tableau.

#### **Step-1: Show With Python**

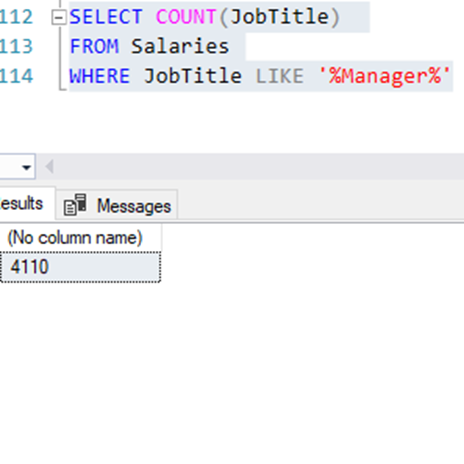
## 

****

**Table 7:People Have the Word “manager” in TheirJob Title with Python**

#### **Step-2: Show with SQL**

## 

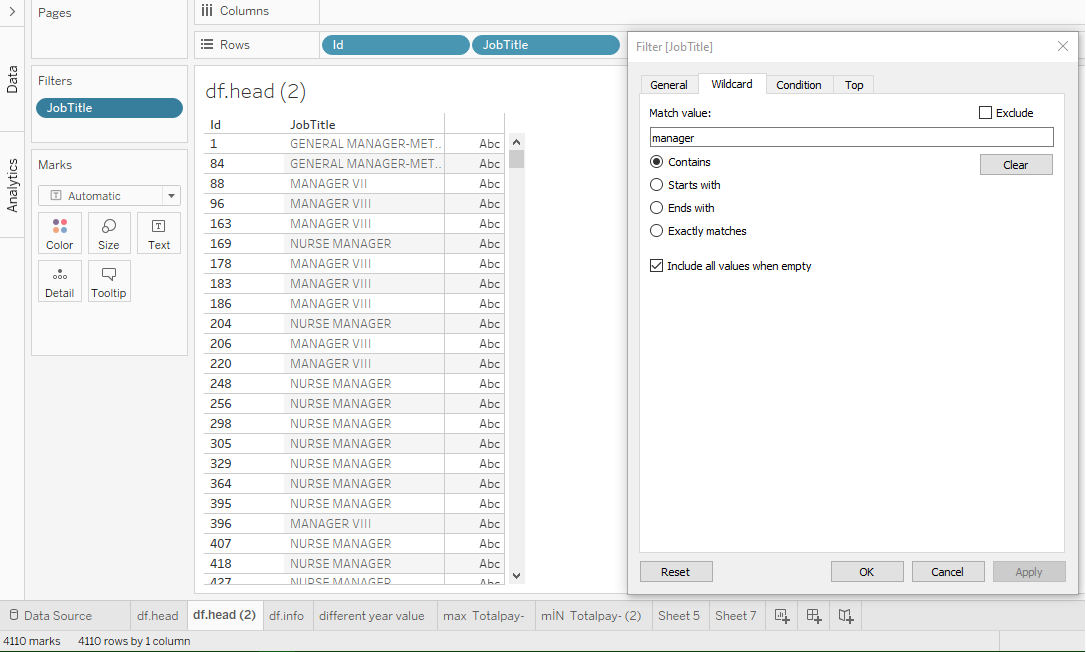
****

**Table 8:People Have the Word “manager” in TheirJob Title with SQL**

#### 

#### **Step 3: Show With Tableau**

## 

****

**Table 9:People Have the Word “manager” in TheirJob Title with Tableau**

Again, we see that there are 4110 records that contain “manager” for all the job titles for Pandas, SQL, and Tableau.

## **Conclusion**

In this article, we analyzed two questions and checked their truthiness with Pandas, SQL, and Tableau. We concurrently used Python, SQL, and Tableau to analyze the data and produced the same outcomes. Thus, we are certain that our analysis is accurate. Using this strategy, advised elbow, you can analyze any dataset with confidence in your analysis's outcomes. I sincerely hope you enjoyed reading. I invite you to embark with me on a learning adventure in data science! If you want to support my writing, keep up with fascinating new practical data science articles by following my Medium page and visiting my GitHub. I appreciate you reading the blog.