**BDAT 1000 Wednesday Data Manipulation Techniques** **Name**: Shubham Chawla

*Assignment 3* **Student ID**: 200493036

* **Used SQL Server Integration Services (SSIS) to insert data from Excel Source “survey\_results\_public” to OLE DB Destination “Stackoverflow” table in Assignment\_3 Database.**
  + **Converted NA to NULL in age column in excel so that MSSQL detects them as numeric columns.**
  + **64,461 rows were inserted into Stackoverflow table in database Assignment\_3**

**Graphical user interface

Description automatically generated**

**An overview of the data inserted (64461 rows)**

**Graphical user interface, application

Description automatically generated**

**Questions**

1. **How many people responded to the survey?**

Used Count() function

**Graphical user interface, text, application, chat or text message

Description automatically generated**

**64461 people responded to the survey.**

1. **Youngest person’s age using StackOverflow?**

**Graphical user interface, text, application, email

Description automatically generated**

According to the survey data, youngest person’s age using StackOverflow is ‘1’ which is quite unusual, this could be a common issue with surveys, responses may include invalid values due to accidental or intentional errors while answering the survey questions. This could be fixed by applying constraints to the age column for example, ignoring any values less than 10 or more than 100 (*WHERE AGE >= 10 or AGE <= 100*). The more likely answer to this question should be 10.

1. **How many people are satisfied with their job?**

**Graphical user interface, text, application

Description automatically generated**

The people who are satisfied with the job either responded as ‘slightly satisfied’ or ‘very satisfied’. The answer is 28,541.

1. **What are top 5 used languages by people?**

**Graphical user interface, text, application

Description automatically generated**

Used String\_Split () function to delimit LanguageWorkedWith Column by ‘;’ delimiter and used cross apply operator to join table valued function (String\_Split ()) with Stackoverflow table.

1. **How many people work as a developer only?**

**Graphical user interface, text, application, Word

Description automatically generated**

Answer: 47,193

1. **Which operating system is used by most of the people?**

**Graphical user interface, text, application, Word

Description automatically generated**

Windows was the most used operating system with 27,170 users.

1. **What are the maximum TOTAL years of code done be any person?**

**Graphical user interface, text, application

Description automatically generated**

YearCode column contains numeric and string values, however the string value ‘More than 50 years’ does not specify the exact number of years therefore I have used ALTER TABLE command to change the data type of ‘YearsCode’ column from String to INT, ignoring the string values.

Maximum total years of code done by any person is 50 years.

**More Queries on this dataset**

1. **List Top 5 Web Frameworks used by people according to the survey.**

**Graphical user interface, text, application

Description automatically generated**

The most used web-framework was jQuery.

1. **What is the education background of all developers?**

**Graphical user interface, text, application

Description automatically generated**

Used update and set command to replace the symbols. We can see that most of the developers were with at least a Bachelors Degree.

1. **What is the country breakdown of the survey respondents with a share of more than 1.5%?**

**Graphical user interface, text, application

Description automatically generated**

United States had around 19.34% of the total survey respondents.

1. **What is the distribution of gender by average work hours?**

**Graphical user interface, text, application

Description automatically generated**

Average working hours for Man were 41, whereas for women and Non-binary, genderqueer, or gender non-conforming they were 40 and 39 respectively