**Ad Click Prediction SQL Project**

**Data overview**

Key columns in the Ad Click Prediction table help the analysis of ad performance. The ID column uniquely identifies every ad impression. Full name is a handy way to personalize ads by representing the individual who viewed them. Gender and age offer demographic information about how various audiences interact with advertisements. Ad position indicates where the advertisement is displayed on the webpage (e.g., Top, Sidebar), while device type logs the type of device (e.g., Mobile, Desktop). The time-of-day logs, when the advertisement was viewed while browsing history, record user preferences based on previous browsing. Whether or not the ad was clicked is indicated in the target column, click.  
  
**objective**

* Cleaned and processed data, handling missing values and fixing inconsistencies (e.g., incorrect gender values).
* Performed data exploration to identify patterns between browsing history and ad clicks.
* Implemented SQL queries to predict the likelihood of ad clicks based on various factors.
* Created triggers for automated data handling and real-time updates.

**Data analysis**

**1-Data Cleaning and Imputation**

A close up of text

Description automatically generatedData Cleaning and Imputation Selected and reviewed records where gender, device type, ad position, browsing history, and time of day were NULL. Imputed missing values for gender by filling in the most frequent value. device type and ad position with the most frequent values. browsing history with a placeholder "No history." time\_\of day with the most frequent value. Created a backup table to store a copy of the cleaned data.

A screenshot of a computer program

Description automatically generated

**2-Data Exploration and Summary Statistics**

* In the Exploratory Data Analysis (EDA), key columns were analysed to understand patterns and distributions. For the Age column, the minimum, maximum, and average values were calculated to assess the range and distribution of user ages.
* The Gender, Device Type, and Ad Position columns were analysed by counting occurrences of each value to observe how these attributes were distributed among users.
* The Click column was examined to calculate click-through rates and identify any engagement patterns. Additionally, the distribution of age, gender, and device type was explored to detect trends, potential outliers, and their impact on the prediction model.

A screenshot of a computer program

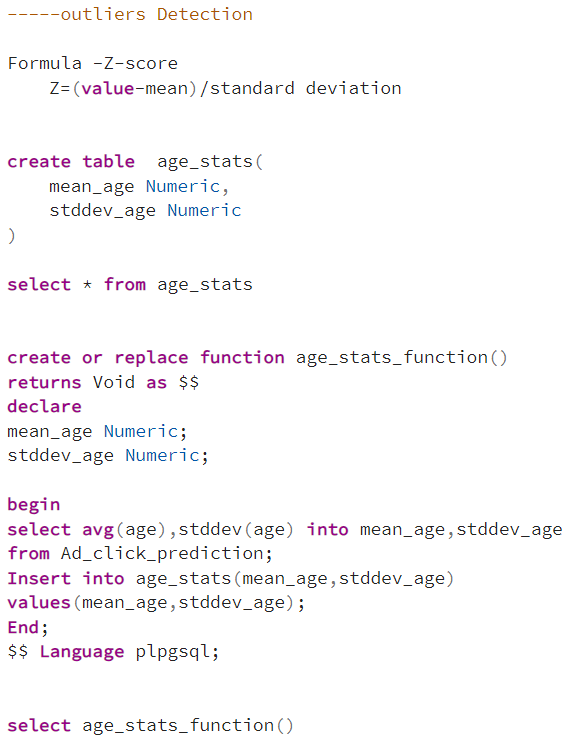
Description automatically generated

**3-Outlier Detection**

* Calculated the Z-score for age using the formula

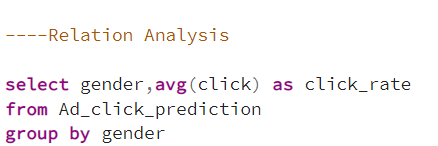
(value - mean)   
 standard deviation

* This helped identify potential outliers, focusing on data points where the Z-score was less than -3 or greater than 3, indicating values significantly different from the average  
  However, no Z-scores were found outside the typical range of -3 to 3, indicating no significant outliers in the dataset based on this method.



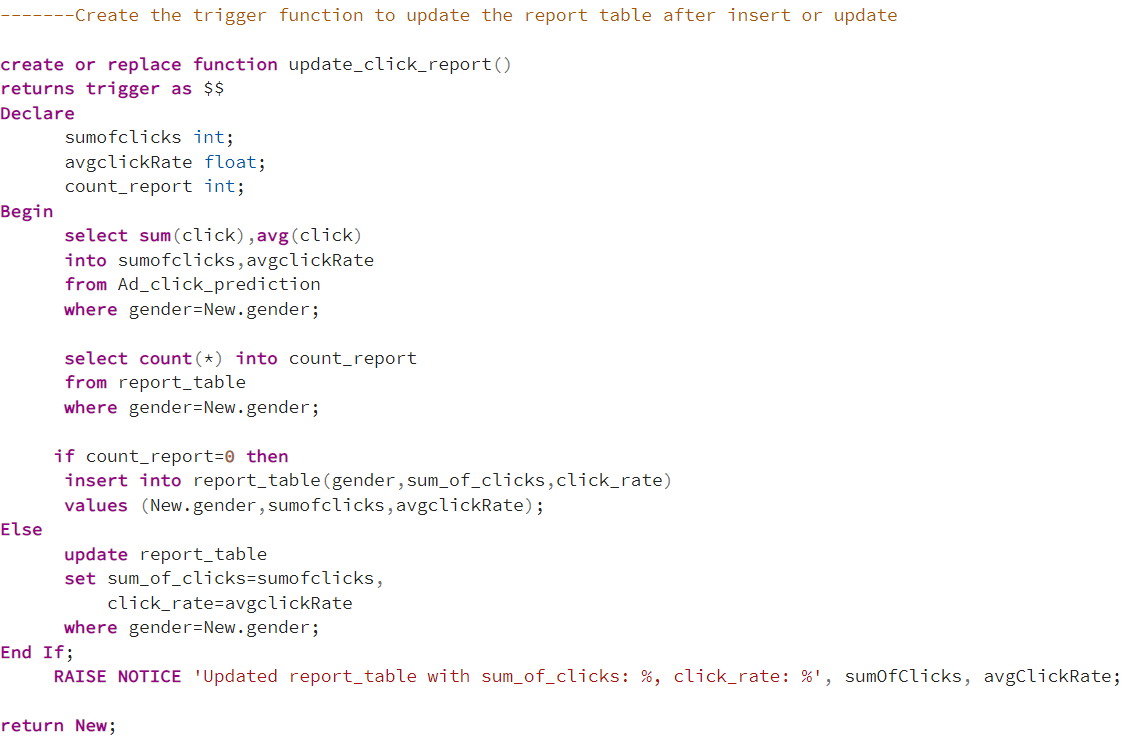
**4- Relation Analysis**

**Analysis Objective:** The main goal is to look into how click-through rates on ads vary by gender in order to better understand audience engagement levels.  
  
**Data Grouping:** For a clear comparison of engagement metrics, we used SQL to group the data by gender and compute the average click rate for each gender segment.  
  
**Trend Identification:** By analysing the data, we can determine which gender has a greater average click-through rate, suggesting possible ad engagement preferences that we can take into account when creating our strategy.  
  
**Targeted Strategy Development:** By creating customized advertisements that appeal to the gender with greater click-through rates, we can improve engagement by modifying ad content, placement, and targeting tactics based on the analysis results.



**5-Report Generation**

Trigger Function was created to record changes to the click column in the ad click data and automate updates to the report table.  
  
**Update click Report table**- Every time a fresh record is added or modified to the Ad click prediction database, a trigger function is created to update the report table. The function updates the values in the report table by calculating the total clicks and clicks rate for each gender. It creates a new record if the gender is absent from the report; if it is present, it modifies the current record.

**Log Click Changes:**  
To save updates to the clicked column, another trigger function was developed. It keeps track of both the old and new click values, recording the modifications and the update timestamp in a click log database. This makes it possible to track user changes and click data in great detail. These triggers guarantee that, regardless of changes in ad click activity, your logs and reports are always updated.