

Cyclistic Data Analytics Process

This document details the data analytics process, including the steps involved and the Excel formulas used for each step.

- **Calculate Trip Duration:** Calculate the duration of each trip by subtracting the start time from the stop time in Excel, creating a new column for enhanced insights.

Formula: =C2-B2 format in cell N2.

- **Calculate Day of Week:** Assigned the day of the week for each trip using the WEEKDAY function in Excel, enhancing temporal analysis.

Formula: =WEEKDAY(C2,1) The 1 argument specifies that Sunday is the first day of the week.

Data Analysis

- **User Type Table:** Created a user type table distinguishing customers and subscribers using COUNTIF in Excel. This classification aids in understanding the composition of the user base.

Formula: =COUNTIF(Trips!J:J; "Customer") and =COUNTIF(Trips!J:J; "Subscriber").

- **Average Trip Duration:** Computed the average trip duration for both customers and subscribers using AVERAGEIFS in Excel, providing insights into user behavior.

Formula: =AVERAGEIFS(Trips!N:N; Trips!J:J; "Customer") and =AVERAGEIFS(Trips!N:N; Trips!J:J; "Subscriber").

- **Group by Age:** Grouped users by age ranges (18-28, 29-38, ..., 79-88) using COUNTIFS in Excel, allowing for targeted analysis of different age demographics.

Formula: =COUNTIFS(Trips!J:J; "Subscriber"; Trips!M:M; ">=18"; Trips!M:M; "<=28").

- Trip Duration by User Type: Generated a table detailing the distribution of trip durations for both customers and subscribers, enabling a comprehensive understanding of ride lengths.

Formula: =COUNTIFS(Trips!N:N; G8:G1896; Trips!J:J; "Customer") and =COUNTIFS(Trips!N:N; G8:G1896; Trips!J:J; "Subscriber").

- Count Routes by User Type: Conducted a detailed analysis of starting and destination stations, creating a count_routes table with user type classifications, station details, and trip counts. Utilized various Excel functions for efficient data extraction and categorization.

The formulas use complex logic to match trip IDs (likely in column A) with route information in a separate sheet named "Route Map".

IF, COUNTIF, MATCH, INDEX, and IFERROR functions are used to achieve this matching and data retrieval based on user type (Subscriber or Customer identified in column I of "Route Map").

Formula:

1. usertype: =IF(COUNTIFS('Route Map'!\$C\$2:\$C\$100, \$A2, 'Route Map'!\$F\$2:\$F\$100, \$D2, 'Route Map'!\$I\$2:\$I\$100, 1) > 0, "Subscriber", "Customer").
2. "from_station_name" Column: =IFERROR(INDEX('Route Map'!\$C\$2:\$C\$100, MATCH(0, COUNTIF(\$A\$1:A1, 'Route Map'!\$C\$2:\$C\$100) + IF('Route Map'!\$I\$2:\$I\$100<>"Subscriber", 1, 0), 0)), "").
3. "from_station_latitude" Column: =IFERROR(INDEX('Route Map'!\$D\$2:\$D\$100, MATCH(0, COUNTIF(\$B\$1:B1, 'Route Map'!\$D\$2:\$D\$100) + IF('Route Map'!\$I\$2:\$I\$100<>"Subscriber", 1, 0), 0)), "").
4. from_station_longitude" Column: =IFERROR(INDEX('Route Map'!\$E\$2:\$E\$100, MATCH(0, COUNTIF(\$C\$1:C1, 'Route Map'!\$E\$2:\$E\$100) + IF('Route Map'!\$I\$2:\$I\$100<>"Subscriber", 1, 0), 0)), "").

5. "to_station_name" Column: =IFERROR(INDEX('Route Map'!\$F\$2:\$F\$100, MATCH(0, COUNTIF(\$D\$1:D1, 'Route Map'!\$F\$2:\$F\$100) + IF('Route Map'!\$I\$2:\$I\$100<>"Subscriber", 1, 0), 0)), "").
6. "to_station_latitude" Column: =IFERROR(INDEX('Route Map'!\$G\$2:\$G\$100, MATCH(0, COUNTIF(\$E\$1:E1, 'Route Map'!\$G\$2:\$G\$100) + IF('Route Map'!\$I\$2:\$I\$100<>"Subscriber", 1, 0), 0)), "").
7. "to_station_longitude" Column: =IFERROR(INDEX('Route Map'!\$H\$2:\$H\$100, MATCH(0, COUNTIF(\$F\$1:F1, 'Route Map'!\$H\$2:\$H\$100) + IF('Route Map'!\$I\$2:\$I\$100<>"Subscriber", 1, 0), 0)), "").
8. "count" Column: =COUNTIFS('Route Map'!\$C\$2:\$C\$100, \$A2, 'Route Map'!\$I\$2:\$I\$100, "Subscriber")

- Top Ten Routes Identification: Employed pandas in Python to filter and identify the top ten routes based on count_routes data, focusing on both customers and subscribers. The results were exported to a CSV file for further examination.