Zuber Ride-Sharing Data Analysis – Chicago

Project Overview

As a data analyst for **Zuber**, a new ride-sharing company launching in Chicago, this project explores rider behavior, preferences, and external influences on ride frequency. Analyze data from competitors and test a hypothesis about the impact of weather on ride frequency.

Data Source:

A database with info on taxi rides in Chicago:

Neighborhood Table, Cab Table, Weather records Table and Trips Table.

Tools:

SQL server – Data analysis

Data Cleaning and Preparation:

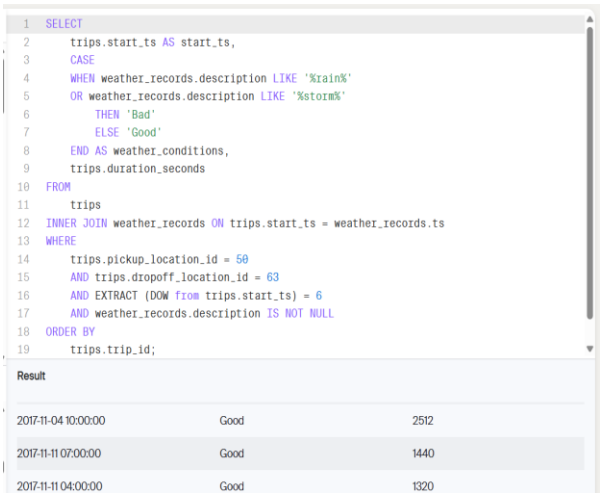
1. Exploratory data analysis.
    - Find the number of taxi rides for each taxi company for November 15-16, 2017, name the resulting field *trips amount.*
    - Find the number of rides for every taxi company whose name contains the words "Yellow" or "Blue" for November 1-7, 2017. Name the resulting variable *trips_amount.* Group the results by the *company_name* field.
    - Join the rides for all other companies in the group "Other." Group the data by taxi company names. Name the field with taxi company names *company*. Sort the result in descending order by *trips_amount*.
    - Retrieve the identifiers of the O'Hare and Loop neighborhoods from the *neighborhoods* table.

2. Investigate whether the duration of rides from the Loop to O'Hare International Airport changes on rainy Saturdays.
- Retrieve the weather condition records from the *weather_records* table. Using the CASE operator, break all hours into two groups: Bad if the *description* field contains the words rain or storm, and Good for others. Name the resulting field *weather_conditions*. The final table must include two fields: date and hour (*ts*) and *weather_conditions*.

- Retrieve from the *trips* table all the rides that started in the Loop (*pickup_location_id:* 50) on a Saturday and ended at O'Hare (*dropoff_location_id*: 63). Get the weather conditions for each ride. Use the method you applied in the previous task. Also, retrieve the duration of each ride. Ignore rides for which data on weather conditions is not available. The table columns should be in the following order: *start_ts, weather_conditions, duration_seconds*
  Sort by *trip_id.*

Data Analysis:

- Retrieve from the *trips* table all the rides that started in the Loop (*pickup_location_id:* 50) on a Saturday and ended at O'Hare (*dropoff_location_id*: 63). Get the weather conditions for each ride. Also, retrieve the duration of each ride.



```
SELECT
    trips.start_ts AS start_ts,
    CASE
    WHEN weather_records.description LIKE '%rain%'
    OR weather_records.description LIKE '%storm%'
        THEN 'Bad'
        ELSE 'Good'
    END AS weather_conditions,
    trips.duration_seconds
FROM
    trips
INNER JOIN weather_records ON trips.start_ts = weather_records.ts
WHERE
    trips.pickup_location_id = 50
    AND trips.dropoff_location_id = 63
    AND EXTRACT (DOW from trips.start_ts) = 6
    AND weather_records.description IS NOT NULL
ORDER BY
    trips.trip_id;
```

Result

2017-11-04 10:00:00	Good	2512
2017-11-11 07:00:00	Good	1440
2017-11-11 04:00:00	Good	1320

Join the rides for all other companies in the group "Other." Group the data by taxi company names. Name the field with taxi company names *company*. Sort the result in descending order by *trips_amount*.

```sql
SELECT
      CASE
     WHEN cabs.company_name IN ('Flash Cab','Taxi Affiliation Services')
         THEN cabs.company_name
         ELSE 'Other'
     END AS company,
   COUNT(trips.trip_id) AS trips_amount
FROM
     trips
INNER JOIN cabs on cabs.cab_id=trips.cab_id
WHERE
     cast(start_ts as date) between '2017-11-01' AND '2017-11-07'
GROUP by
    company
ORDER BY
trips_amount desc;
```

Result

company	trips_amount
Other	335771
Flash Cab	64084

Analysis:

- Investigated no. of taxi rides, Group the results by company name, retrieve the weather condition weather its good or bad and rain or storm.
- Tested the hypothesis that weather conditions significantly impact the number of rides, using integrated weather and ride datasets.
- Insights from this analysis help inform Zuber's operational strategy, pricing models, and targeted marketing based on passenger behavior and seasonal trends.

Recommendations:

Weather-Aware Pricing & Promotions

Implement surge pricing or targeted promotions during rainy, snowy, or extremely cold days—ride demand increases under adverse weather conditions, offering revenue opportunities.

Focus on High-Demand Time Windows

Concentrate driver availability and marketing efforts during weekday commute hours (7–9 AM and 5–7 PM) and weekend evenings, when ride frequency peaks.

Seasonal Marketing Campaigns

Allocate more advertising budget during months with high ride frequency and favorable profit margins—especially during fall and early winter, when weather-driven demand spikes.