

SQL

```
CREATE DATABASE IF NOT EXISTS uber_data;
USE uber_data;
```

```
DROP TABLE IF EXISTS uber_request_data;
DROP TABLE IF EXISTS uber_request_data_temp;
```

```
CREATE TABLE uber_request_data (
    request_id INT PRIMARY KEY,
    pickup_point VARCHAR(50) NOT NULL,
    driver_id INT,
    status VARCHAR(20) NOT NULL,
    request_timestamp DATETIME NOT NULL,
    drop_timestamp DATETIME
);
```

```
CREATE TABLE uber_request_data_temp (
    request_id TEXT,
    pickup_point TEXT,
    driver_id TEXT,
    status TEXT,
    request_timestamp_text TEXT,
    drop_timestamp_text TEXT
);
```

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/Uber
Request Data.csv'
INTO TABLE uber_request_data_temp
FIELDS TERMINATED BY ','
OPTIONALLY ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS
(request_id, pickup_point, driver_id, status, request_timestamp_text,
drop_timestamp_text);
```

```
INSERT INTO uber_request_data (
    request_id,
    pickup_point,
    driver_id,
    status,
    request_timestamp,
    drop_timestamp
)
SELECT
    CAST(TRIM(REPLACE(request_id, '\r', '')) AS UNSIGNED),
    TRIM(REPLACE(pickup_point, '\r', '')),
    CASE
        WHEN TRIM(REPLACE(driver_id, '\r', '')) IN ('NA', '') THEN NULL
        ELSE CAST(TRIM(REPLACE(driver_id, '\r', '')) AS UNSIGNED)
    END,
    TRIM(REPLACE(status, '\r', '')),
    CASE
        WHEN TRIM(REPLACE(request_timestamp_text, '\r', '')) IN ('NA',
'' ) THEN NULL
```

```

        ELSE STR_TO_DATE(TRIM(REPLACE(request_timestamp_text, '\r', '')),
'%d-%m-%Y %H:%i:%s')
    END,
    CASE
        WHEN TRIM(REPLACE(drop_timestamp_text, '\r', '')) IN ('NA', '')
THEN NULL
        ELSE STR_TO_DATE(TRIM(REPLACE(drop_timestamp_text, '\r', '')),
'%d-%m-%Y %H:%i:%s')
    END
FROM uber_request_data_temp;

```

```

-- Total number of rides
SELECT COUNT(*) FROM uber_request_data;

```

```

-- Rides per status
SELECT status, COUNT(*) FROM uber_request_data GROUP BY status;

```

```

-- Average trip duration (in minutes)
SELECT
    AVG(TIMESTAMPDIFF(MINUTE, request_timestamp, drop_timestamp)) AS
avg_trip_duration
FROM uber_request_data
WHERE drop_timestamp IS NOT NULL;

```

```

-- Unique drivers
SELECT COUNT(DISTINCT driver_id) AS total_drivers FROM uber_request_data;

```

```

-- Distinct status types
SELECT DISTINCT status FROM uber_request_data;

```

```

-- Cancelled rides
SELECT * FROM uber_request_data WHERE status = 'cancelled';

```