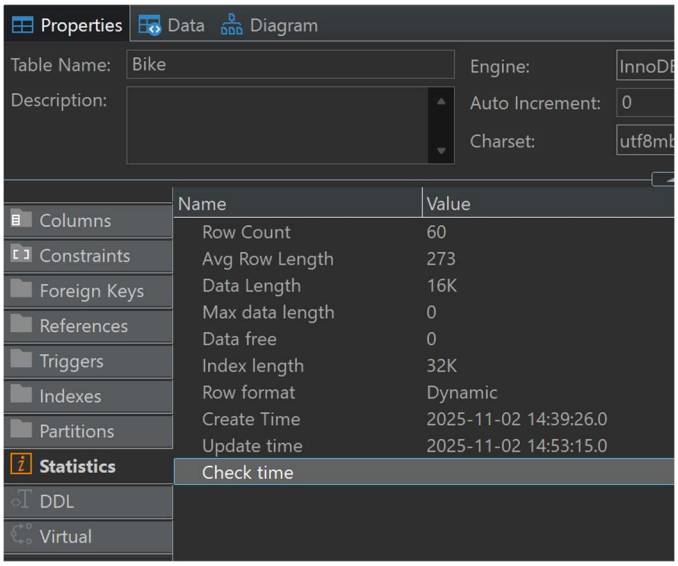
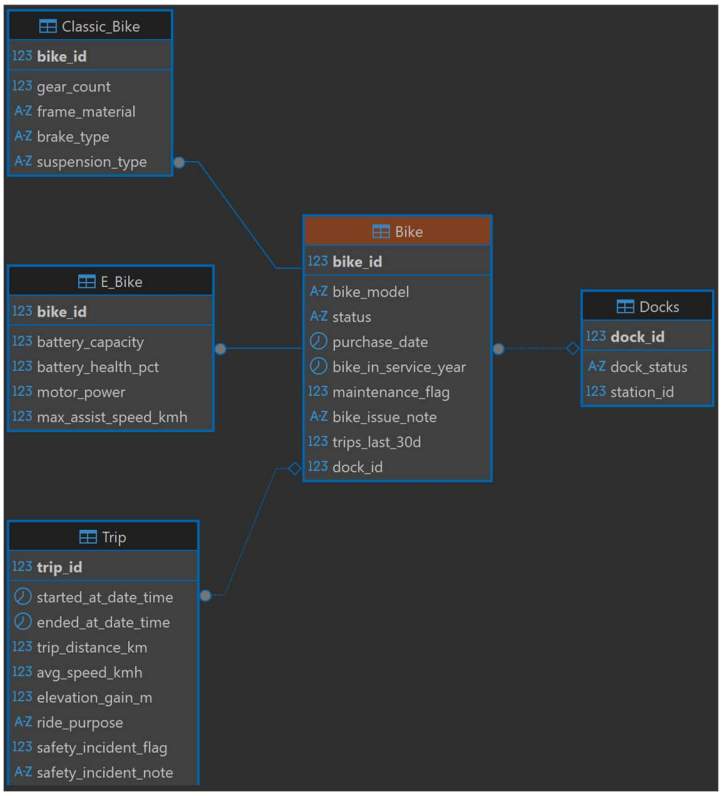
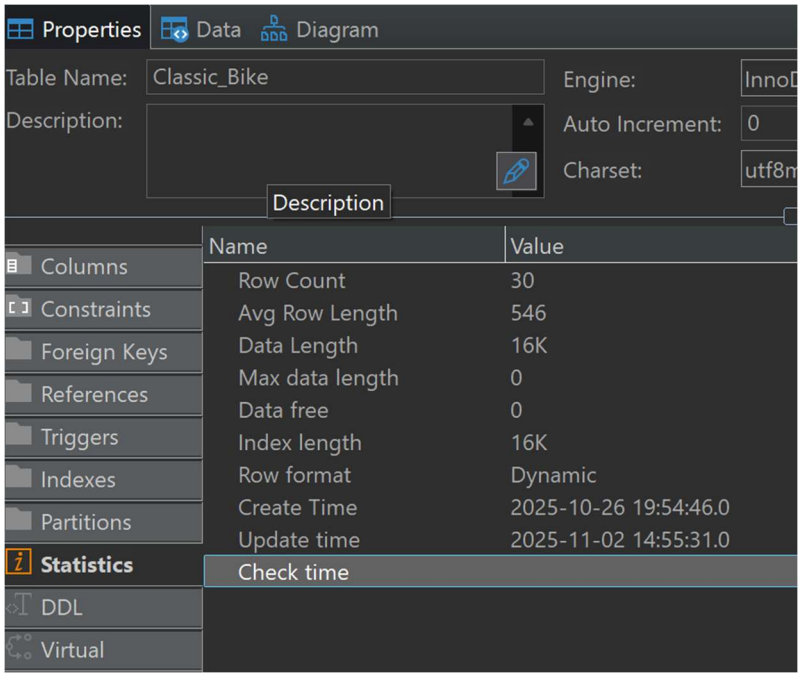
Project Deliverable 5 – Table Data Populations
Group 3 - Sabereen Safa, Jordan Eberhart, James Eastwood
The work contained here is the work of Group 3 and Group 3 alone.



Bike_ID was synthesized using AI as a sequential number with the leading 32 standing for Classic Bike and the leading 52 standing for Electric Bike, then zeroes before the sequence starting at 1.

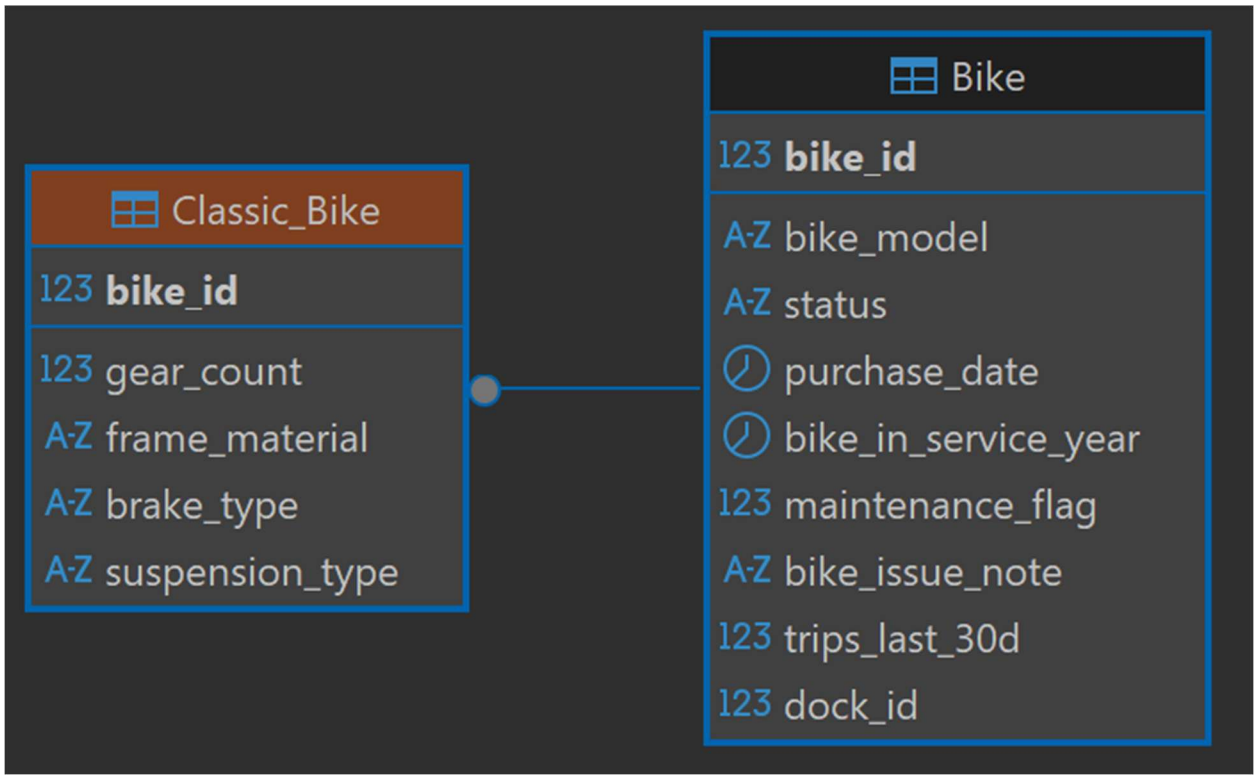
Purchase_Date was synthesized data using AI based on the provided data of Bike_In_Service_Year.

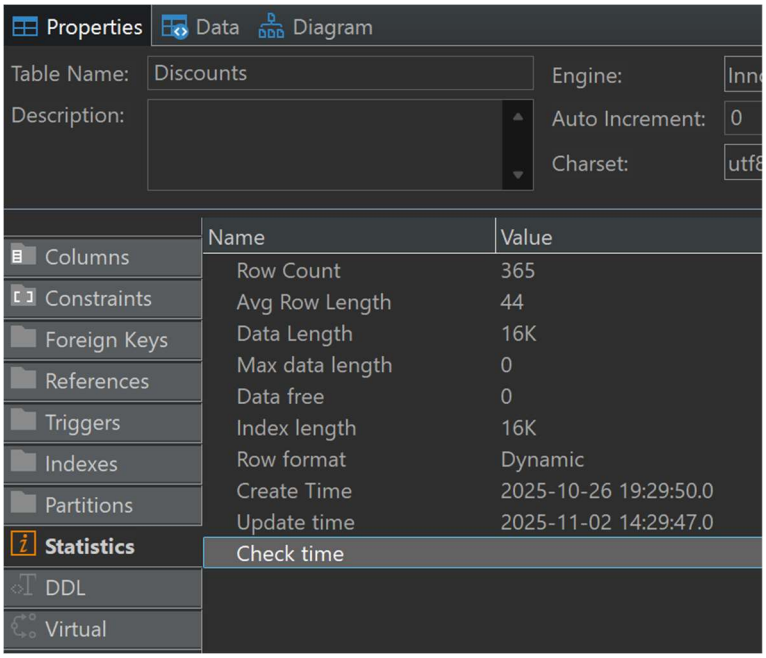




Gear_count was synthesized using AI falling within the constraint of the 25 gear max.

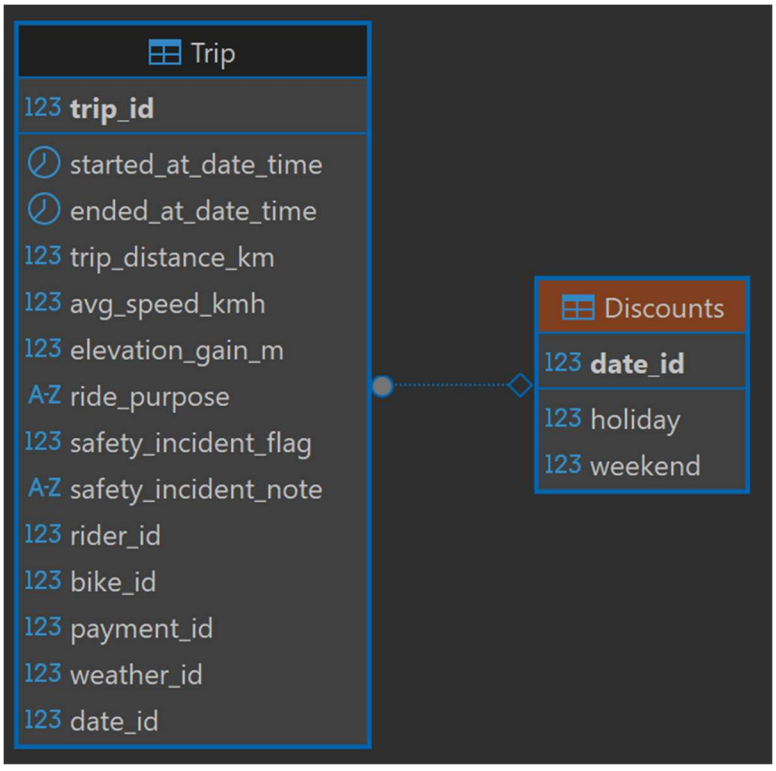
Frame_material, brake_type, suspension_type, were also synthesized using AI and randomly assigned to bike_ids.

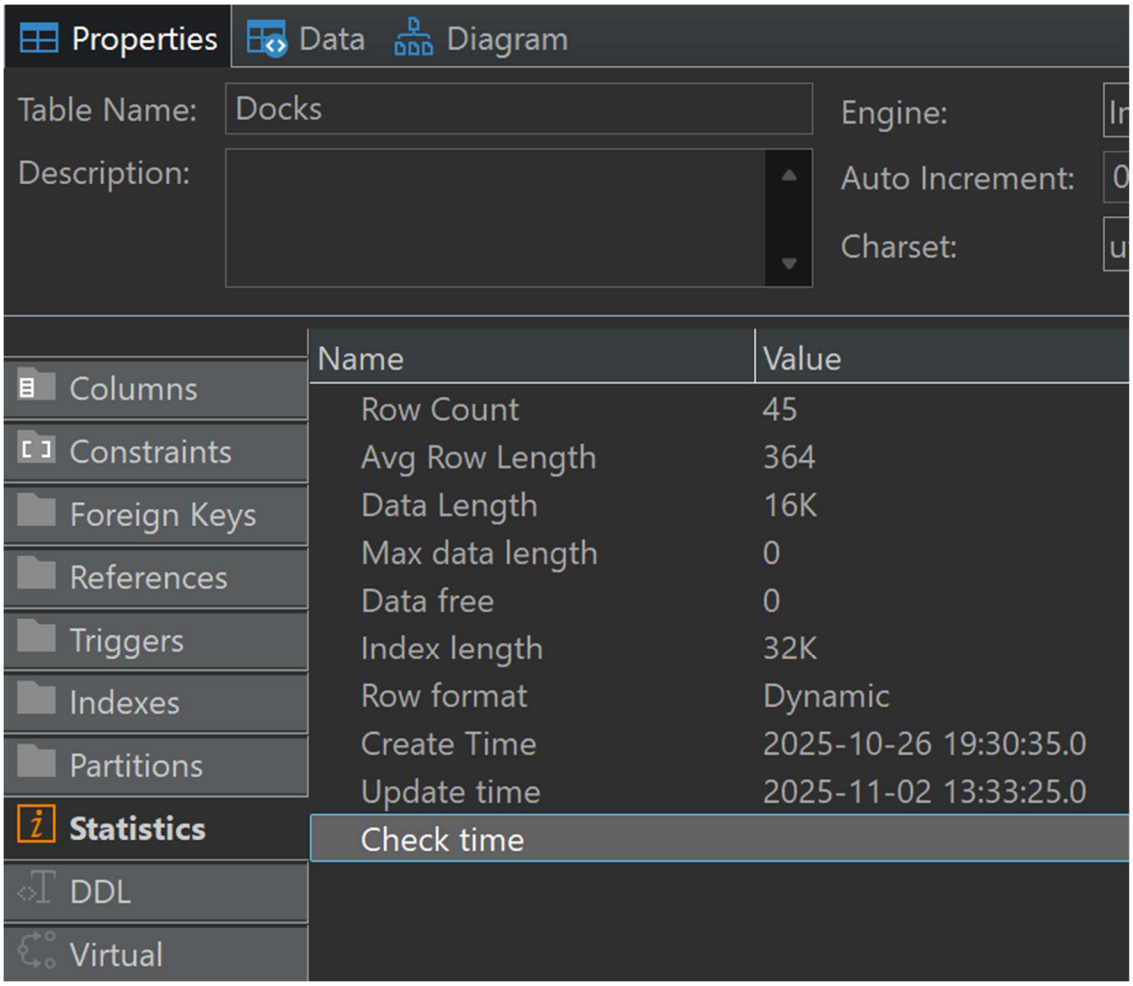




Date_id was synthesized with 49193 standing for DISC because of the name of the Table "Discounts." Then preceding zeroes and then a sequential number starting at 1 and ending at 366 (since 2024 was a Leap Year) were added.

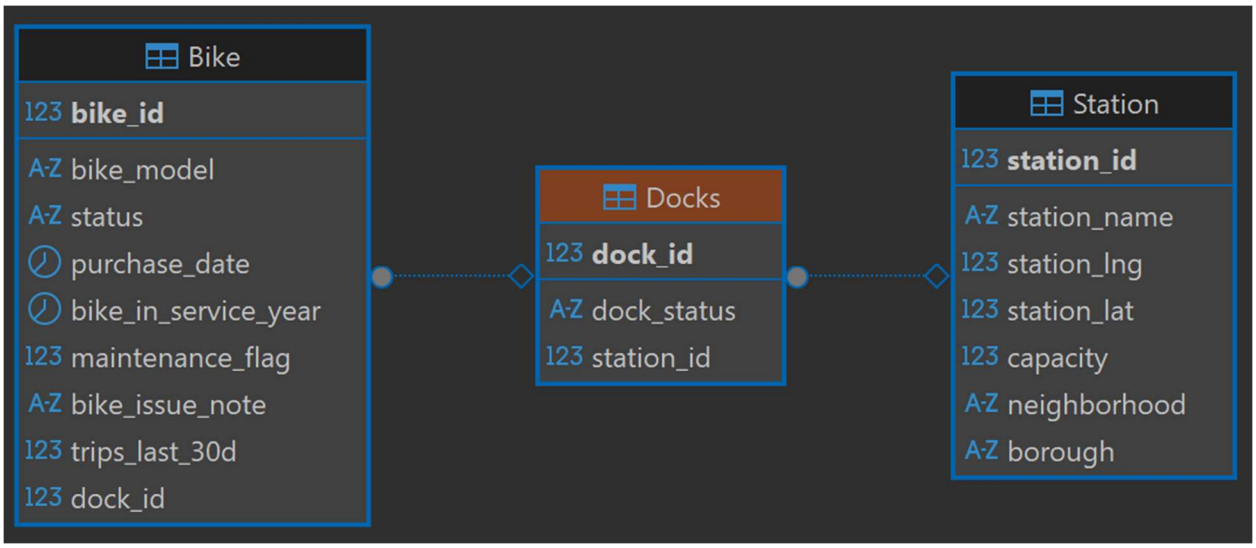
Holiday and Weekend are boolean variables which were synthesized based on whether they were holidays or weekends in the year of 2024.

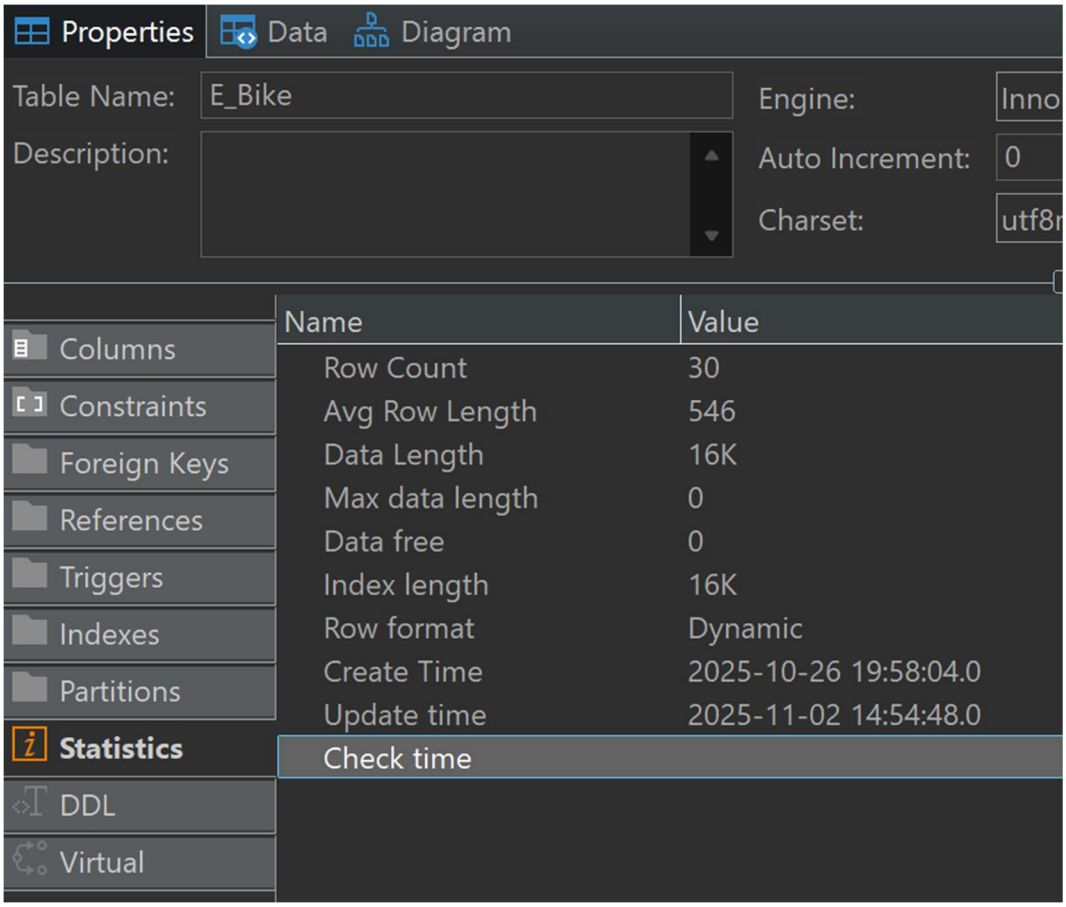




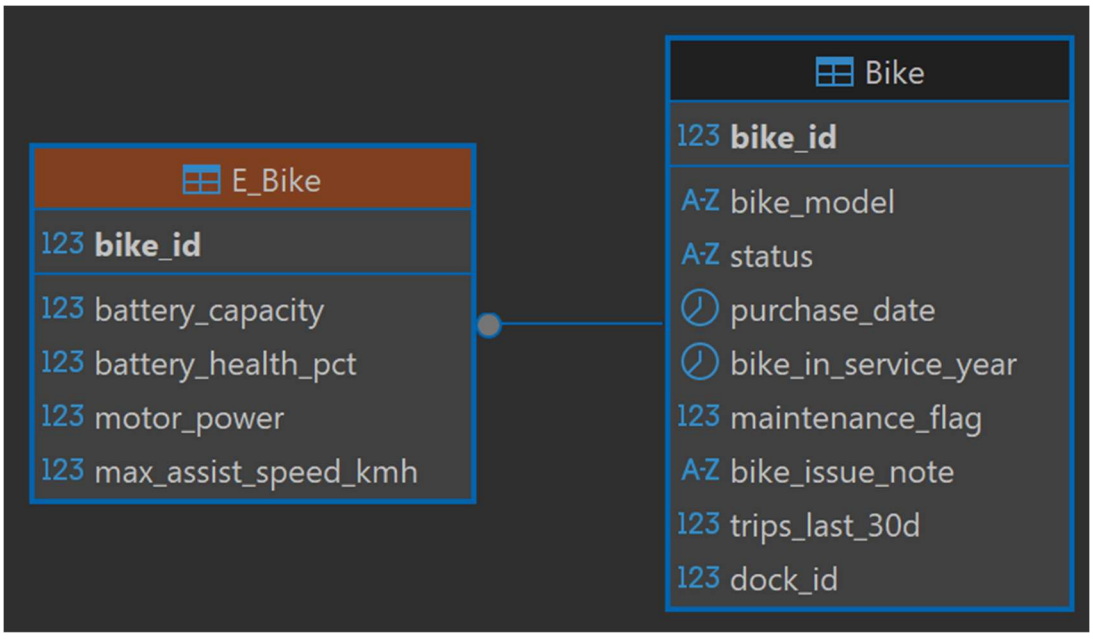
Dock_id was an AI synthesized sequential number with a leading 4 representing “D” for Dock and then zeroes to separate the actual dock_id.

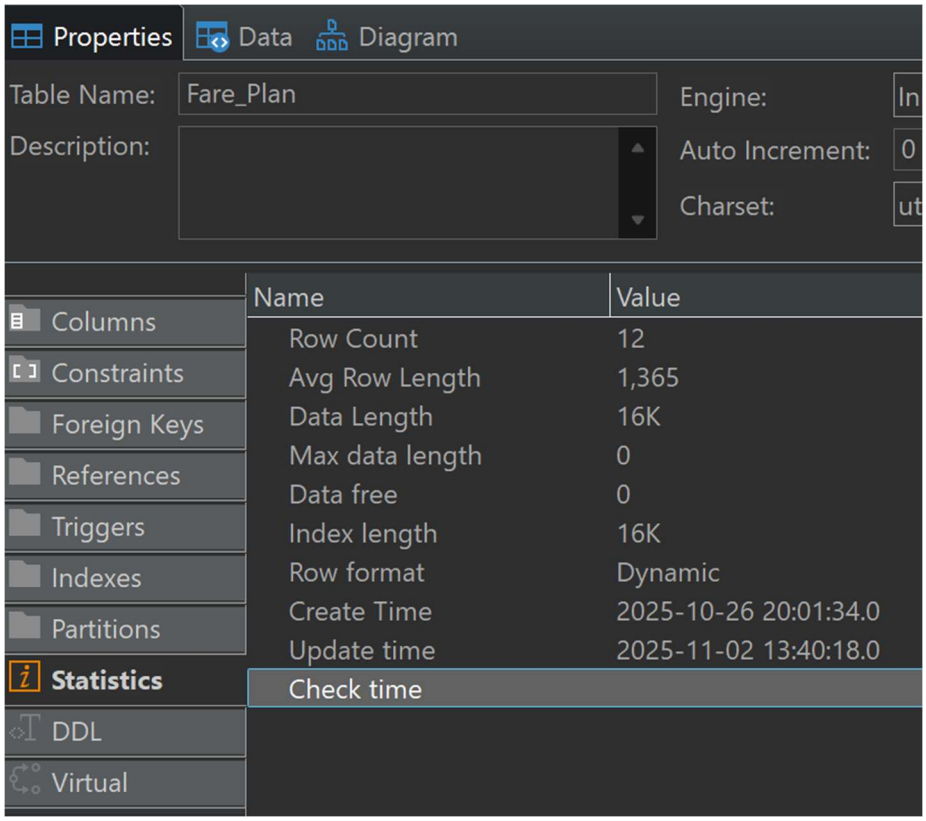
Dock_status was randomly selected between “available,” “occupied,” or “out of service.”





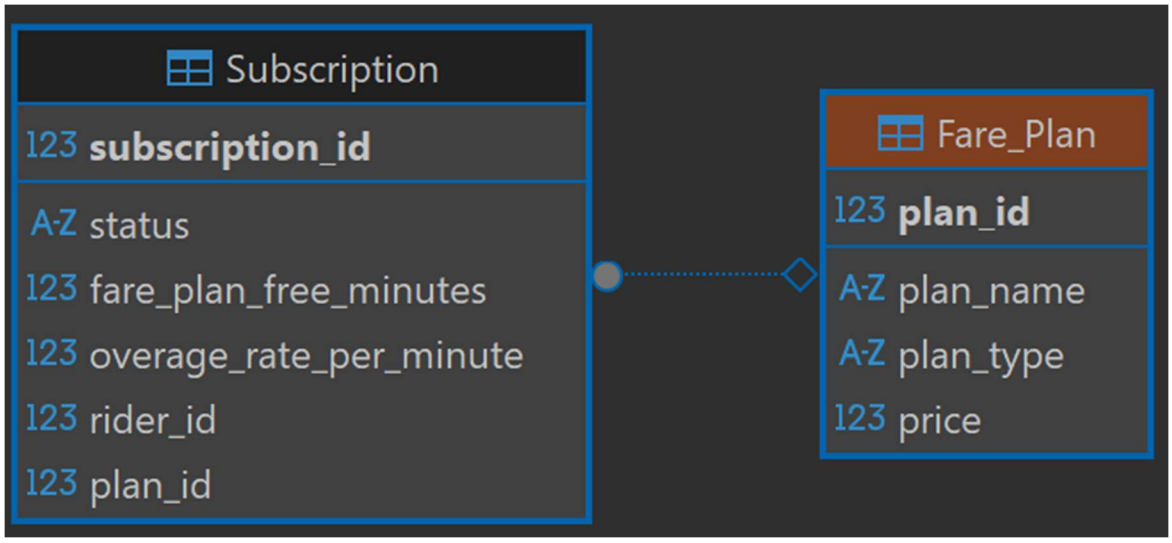
Battery_capacity, battery_health_pct, motor_power, and max_assist_speed_kmh were all randomly AI synthesized.

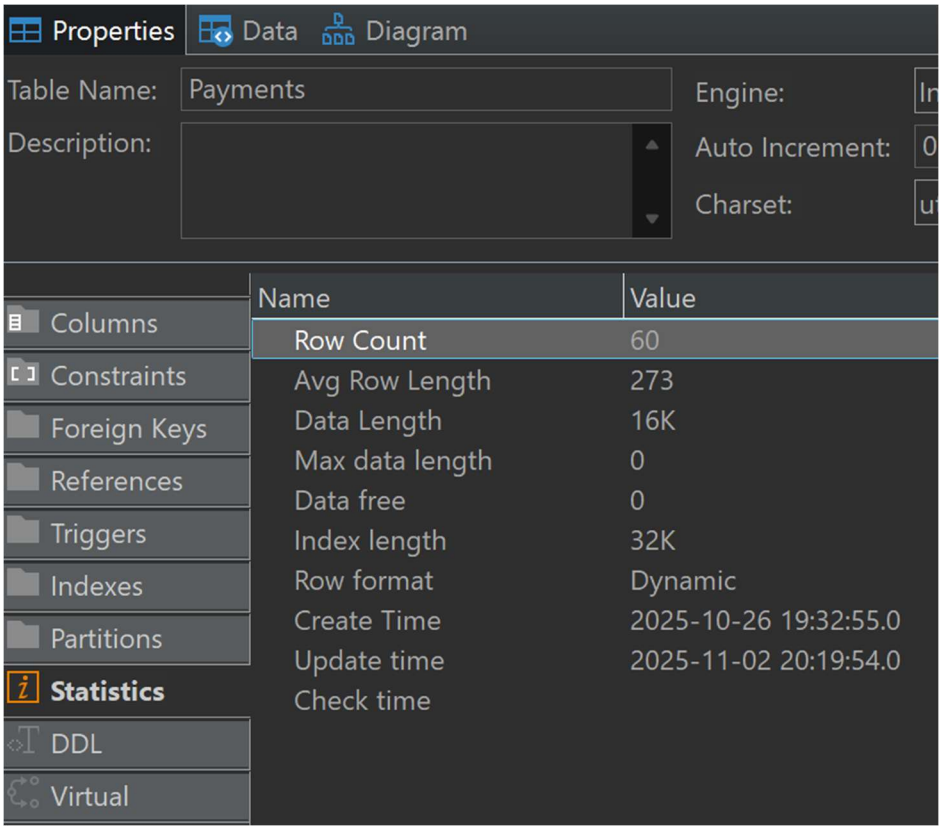




Plan_id was an AI sythensized random number with a lead 616 for “F” and “P” for the Fare_Plan table. Then there were separating zeroes and the sequential number of all of the plans, of which there are only 12 due to there not being any additional reasonable fare plan options.

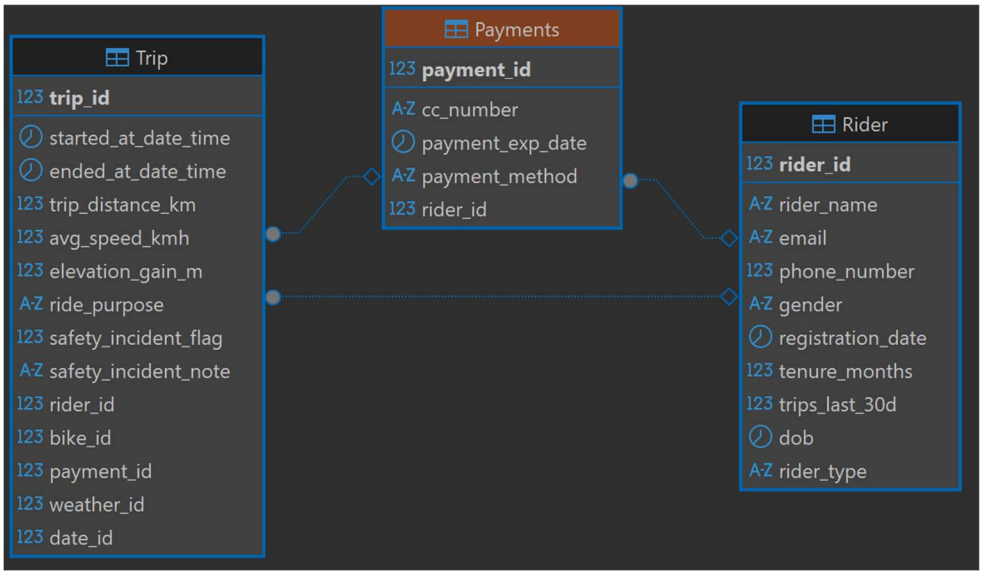
Plan_name, plan_type, and price were all AI generated and chosen as the best options to differentiate between fare plans with data that was reasonable and realistic.

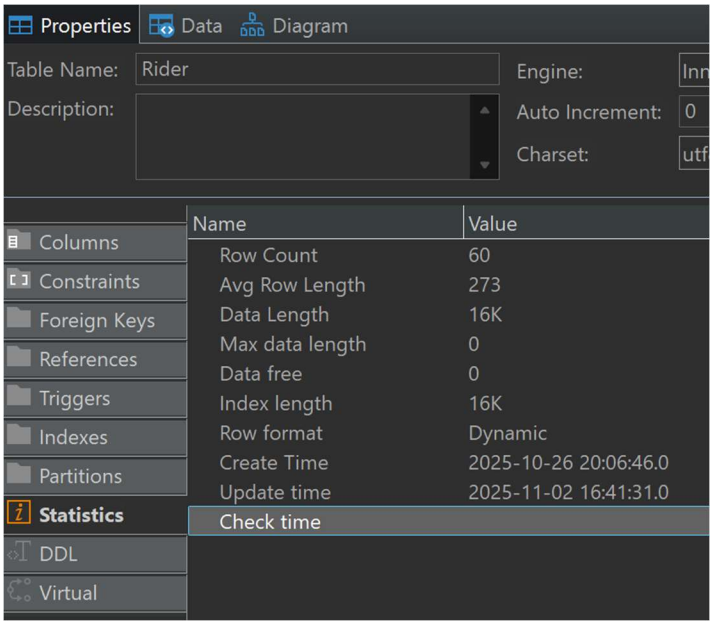




Payment_id was an AI synthesized, random number with a leading 16 for “P” for Payments, then a separator zero, and then the random number for a grand total of 10 digits.

CC_number, payment_exp_date, and payment_method were all AI synthesized numbers where the Visa, Mastercard, and Discover payment_method have 16 digit CC_numbers while American Express have 15 digits. Payment_exp_dates were generated to be 2 to 6 years into the future.





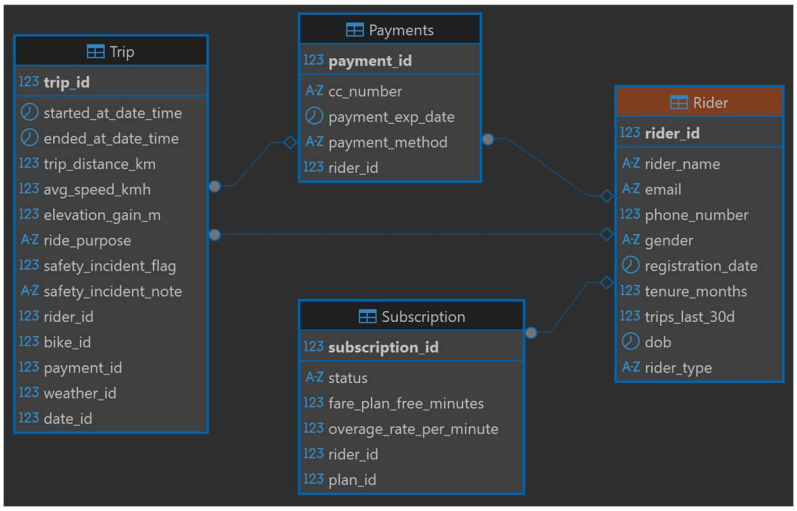
Rider_id was an AI synthesized number with a leading 18 for “R” for Rider, then a separator zero, and then the random number for a grand total of 10 digits.

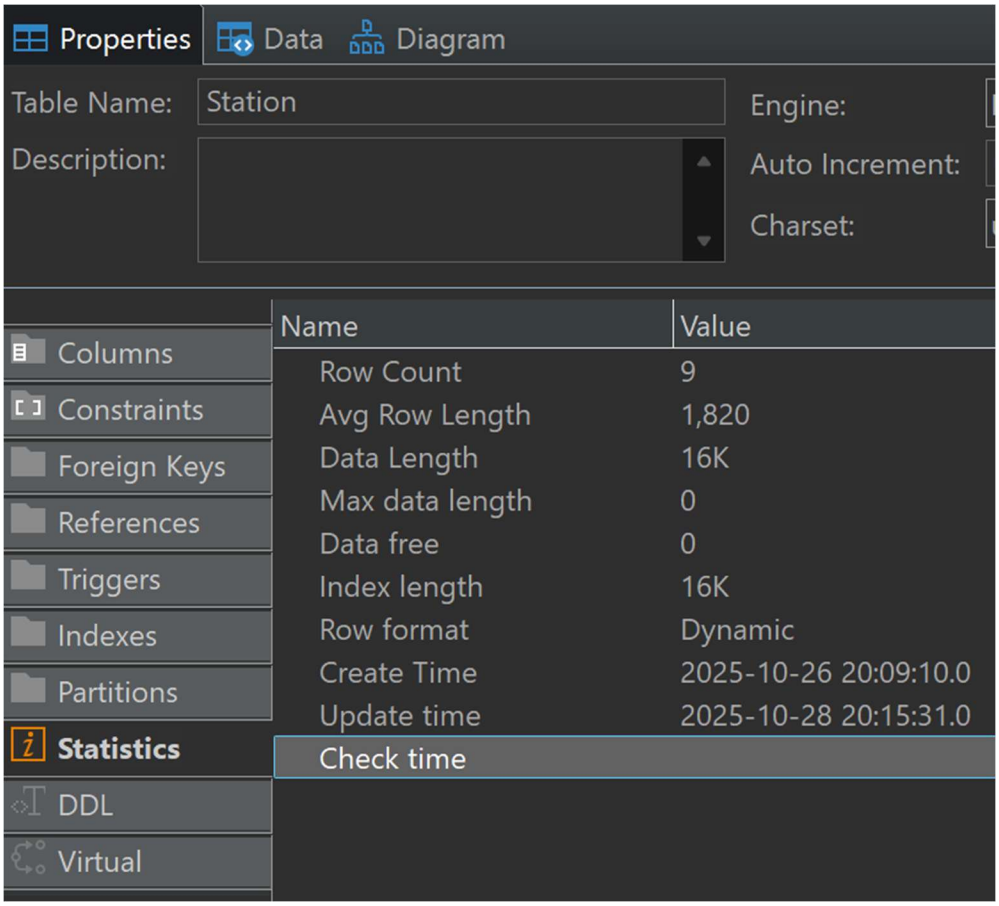
Rider_name, email, phone_number, and dob were all randomly generated by AI.

Gender was randomly chosen from “M,” “F,” “Other,” or “Prefer Not to Say,” with a bias put towards M and F if the name of the rider was a generally M or F name.

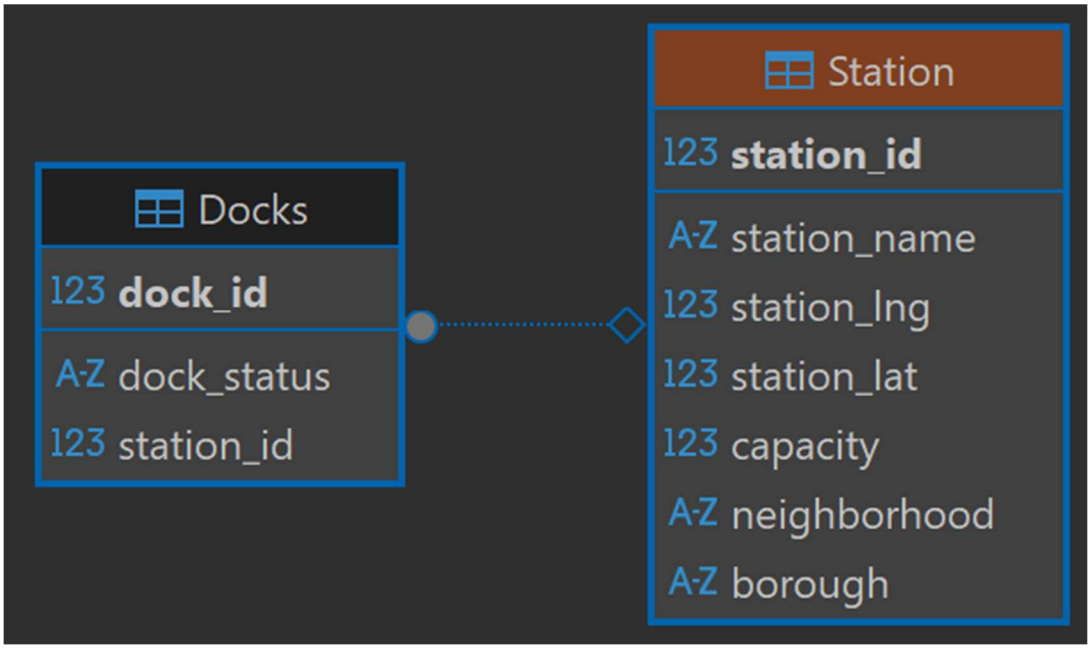
Registration_date was based on tenure_months which was a provided datapoint in the original dataset. If tenure_months was 6, then the registration date would have been a randomly generated date in a 30 day window surrounding the ride date, 6 months in the past.

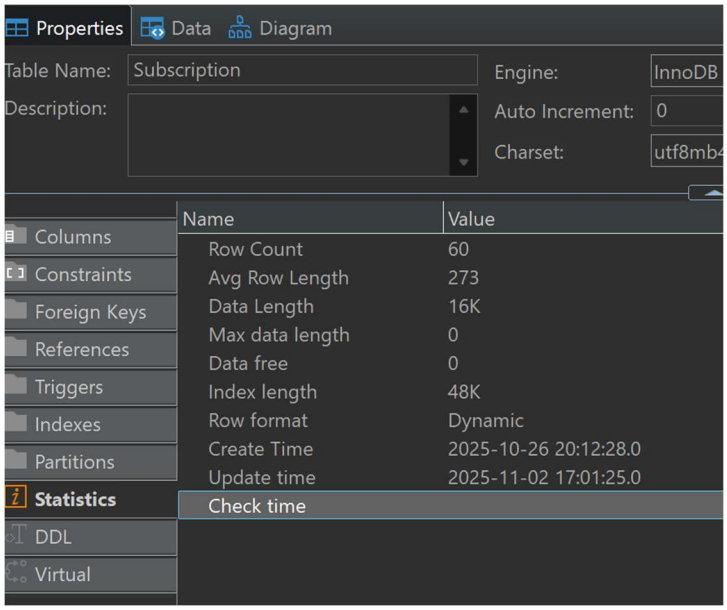
Tenure_months and trips_last_30d were data that was all provided in the original dataset.





Station_id was an ID number provided in the original dataset, along with every other attribute and value. As the only table that needed no data generated, if we just left it alone, even though it only has 9 records, it made sense to leave it untouched.

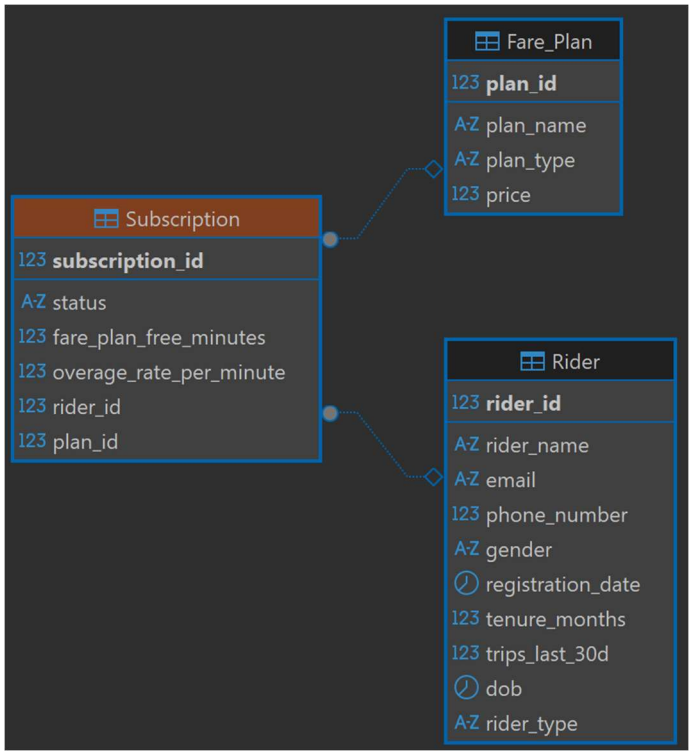


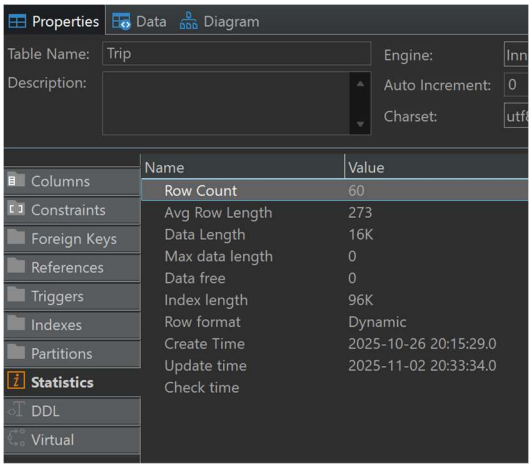


Subscription_id was an AI synthesized, random number. The number is led by 5 digits 19212 for “S,” “U,” “B” for SUB in Subscription, then a separator zero, and then the random number for a total of 10 digits.

Fare_plan_free_minutes, and overage_rate_per_minute are actual data included in the original dataset.

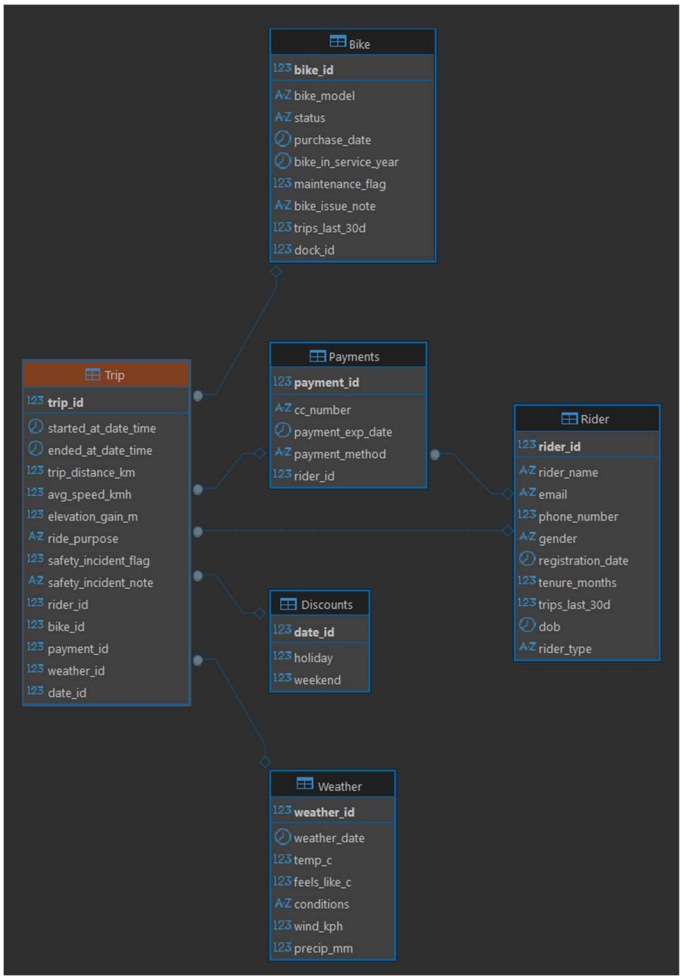
Status was an AI randomly chosen selection between “active,” “cancelled,” and “paused.”

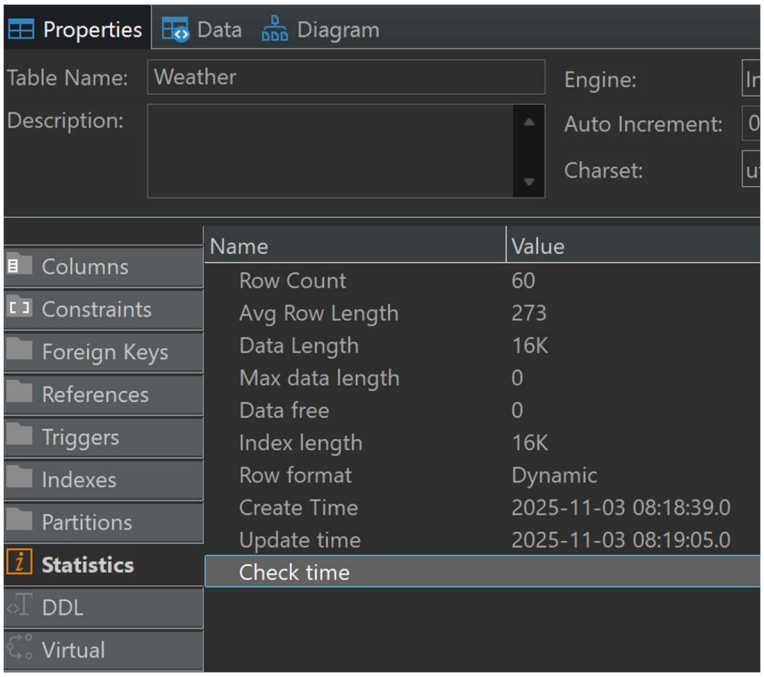




Trip_id was an AI synthesized sequential number with a leading 20 for “T” in Trip, and then a lot of separator zeroes, followed by the sequence starting at 1, with a grand total of 10 digits.

Every other column, aside from the Foreign Keys are data that was provided in the original dataset.





Weather_id is an AI synthesized number led by 2321 for “W,” and “U” because when we first generated the weather_id data, we mistakenly generated a varchar identifier with a leading WU, which I then converted with 2321. Then there were separator zeroes and a sequential number starting at 1 for a grand total of 10 digits.

All other data in the table was provided by the original dataset.