DAMG 7370 - Designing Advanced Data Architectures for Business Intelligence

FINAL PROJECT

Team Members:

Dharun Karthick Ramaraj

Nitin Sai Varma Indukuri

Nikhil Godalla

Linata Deshmukh

Dataset

Motor Vehicle Collisions - Crashes

The Motor Vehicle Collisions crash table contains details on the crash event. Each row represents a crash event. The Motor Vehicle Collisions data tables contain information from all police reported motor vehicle collisions in NYC. The police report (MV104-AN) is required to be filled out for collisions where someone is injured or killed, or where there is at least $1000 worth of damage (https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/ny_overlay_mv-104an_rev05_2004.pdf). It should be noted that the data is preliminary and subject to change when the MV-104AN forms are amended based on revised crash details.For the most accurate, up to date statistics on traffic fatalities, please refer to the NYPD Motor Vehicle Collisions page (updated weekly) or Vision Zero View (updated monthly).

Due to success of the CompStat program, NYPD began to ask how to apply the CompStat principles to other problems. Other than homicides, the fatal incidents with which police have the most contact with the public are fatal traffic collisions. Therefore, in April 1998, the Department implemented TrafficStat, which uses the CompStat model to work towards improving traffic safety. Police officers complete form MV-104AN for all vehicle collisions. The MV-104AN is a New York State form that has all of the details of a traffic collision. Before implementing Trafficstat, there was no uniform traffic safety data collection procedure for all of the NYPD precincts. Therefore, the Police Department implemented the Traffic Accident Management System (TAMS) in July 1999 in order to collect traffic data in a uniform method across the City. TAMS required the precincts manually enter a few selected MV-104AN fields to collect very basic intersection traffic crash statistics which included the number of accidents, injuries and fatalities. As the years progressed, there grew a need for additional traffic data so that more detailed analyses could be conducted. The Citywide traffic safety initiative, Vision Zero started in the year 2014. Vision Zero further emphasized the need for the collection of more traffic data in order to work towards the Vision Zero goal, which is to eliminate traffic fatalities. Therefore, the Department in March 2016 replaced the TAMS with the new Finest Online Records Management System (FORMS). FORMS enables the police officers to electronically, using a Department cellphone or computer, enter all of the MV-104AN data fields and stores all of the MV-

104AN data fields in the Department's crime data warehouse. Since all of the MV-104AN data fields are now stored for each traffic collision, detailed traffic safety analyses can be conducted as applicable.

Austin Crash Report Data - Crash Level Records

Crash data is obtained from the Texas Department of Transportation (TXDOT) Crash Record Information System (CRIS) database, which is populated by reports submitted by Texas Peace Officers throughout the state, including Austin Police Department (APD), and maintained by TXDOT.

This dataset contains crash-level records for crashes which have occurred in the last ten years. Crash data may take several days or weeks to be initially provided and finalized as it is furnished to the Austin Transportation & Public Works Department, therefore a two-week delay is implemented to help ensure more accurate and complete results.

Please note that the data and information on this website is for informational purposes only. While we seek to provide accurate information, please note that errors may be present and information presented may not be complete.

Traffic Crashes - Crashes

Crash data shows information about each traffic crash on city streets within the City of Chicago limits and under the jurisdiction of Chicago Police Department (CPD). Data are shown as is from the electronic crash reporting system (E-Crash) at CPD, excluding any personally identifiable information. Records are added to the data portal when a crash report is finalized or when amendments are made to an existing report in E-Crash. Data from E-Crash are available for some police districts in 2015, but citywide data are not available until September 2017. About half of all crash reports, mostly minor crashes, are self-reported at the police district by the driver(s) involved and the other half are recorded at the scene by the police officer responding to the crash. Many of the crash parameters, including street condition data, weather condition, and posted speed limits, are recorded by the reporting officer based on best available information at the time, but many of these may disagree with posted information or other assessments on road conditions. If any new or updated information on a crash is received, the reporting officer may amend the crash report at a later time. A traffic crash within the city limits for which CPD is not the responding police agency, typically crashes on interstate highways, freeway ramps, and on local roads along the City

boundary, are excluded from this dataset.

All crashes are recorded as per the format specified in the Traffic Crash Report, SR1050, of the Illinois Department of Transportation. The crash data published on the Chicago data portal mostly follows the data elements in SR1050 form. The current version of the SR1050 instructions manual with detailed information on each data elements is available here.

As per Illinois statute, only crashes with a property damage value of $1,500 or more or involving bodily injury to any person(s) and that happen on a public roadway and that involve at least one moving vehicle, except bike dooring, are considered reportable crashes. However, CPD records every reported traffic crash event, regardless of the statute of limitations, and hence any formal Chicago crash dataset released by Illinois Department of Transportation may not include all the crashes listed here.

Change 11/21/2023: We have removed the RD_NO (Chicago Police Department report number) for privacy reasons.

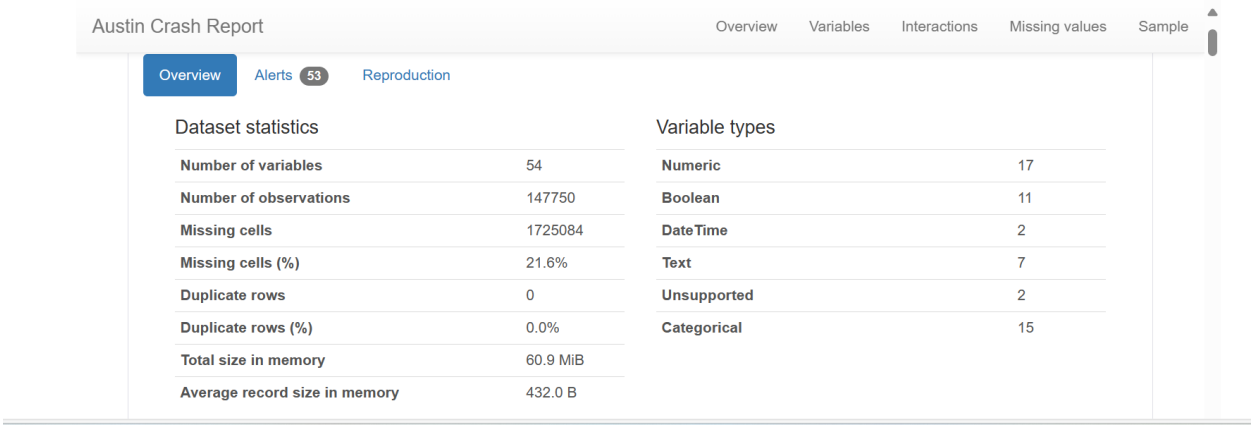
Data Profiling

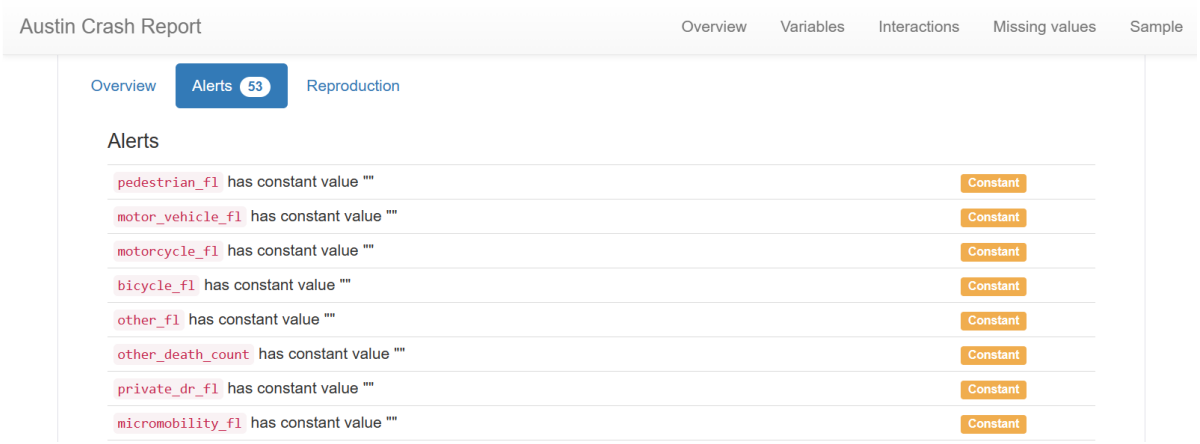
Profiling: ydata_profiling

Austin Crash Report

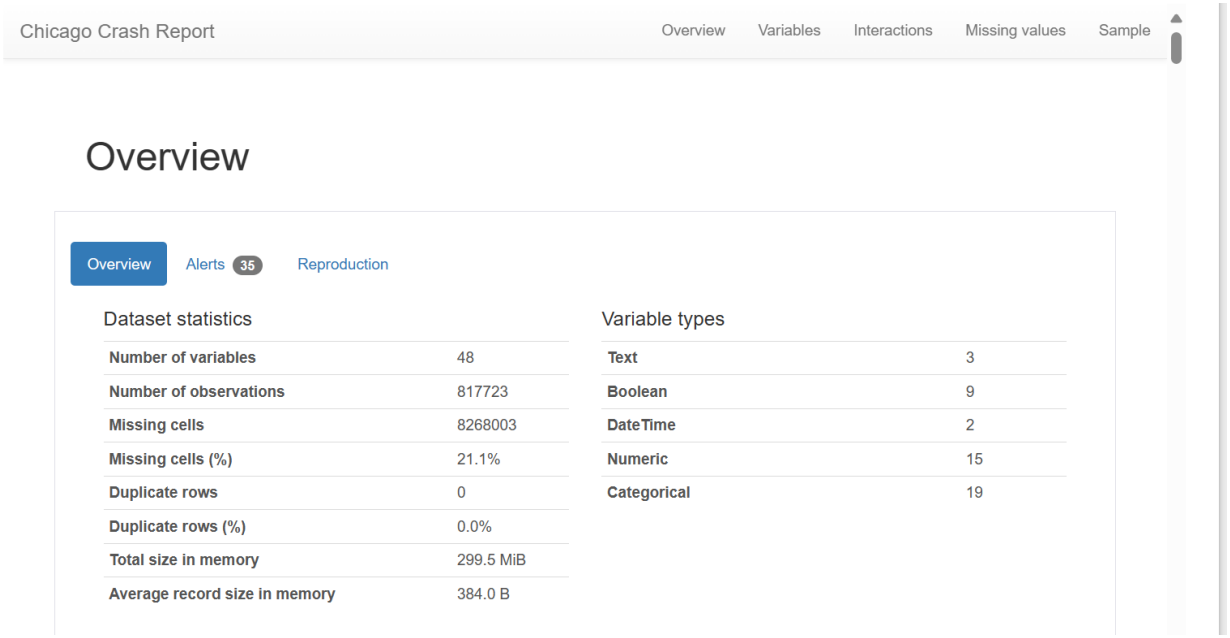
```
[10]: AustinProfile.to_notebook_iframe()
```

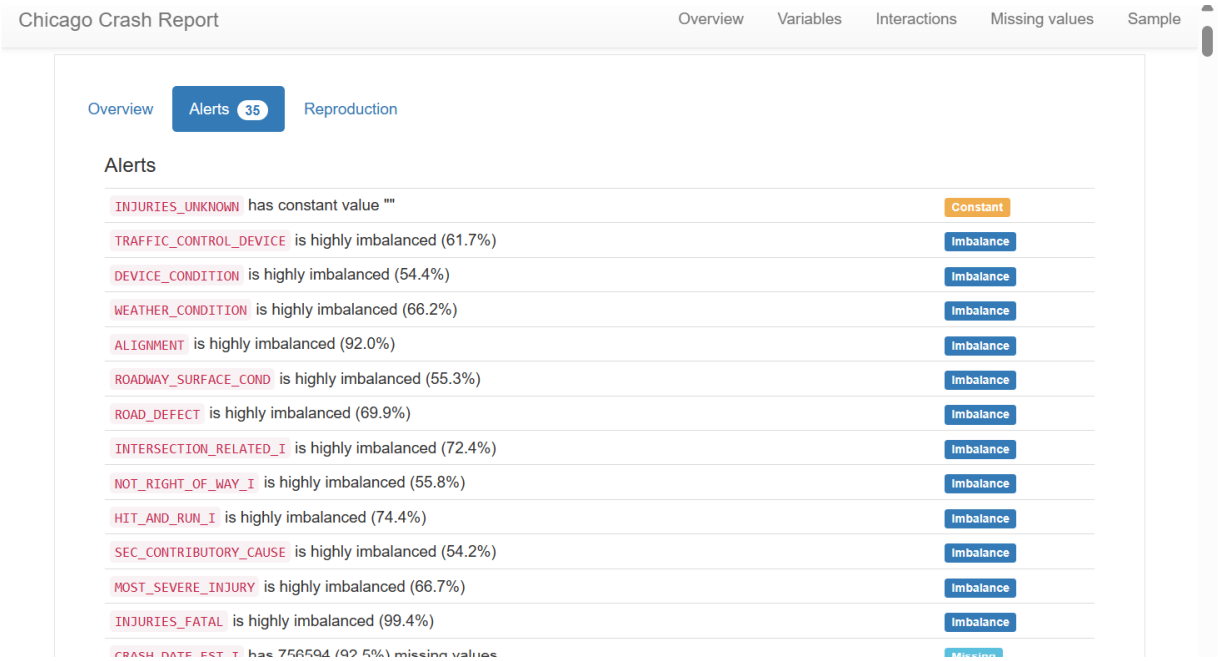
Error displaying widget: model not found
Error displaying widget: model not found
Error displaying widget: model not found



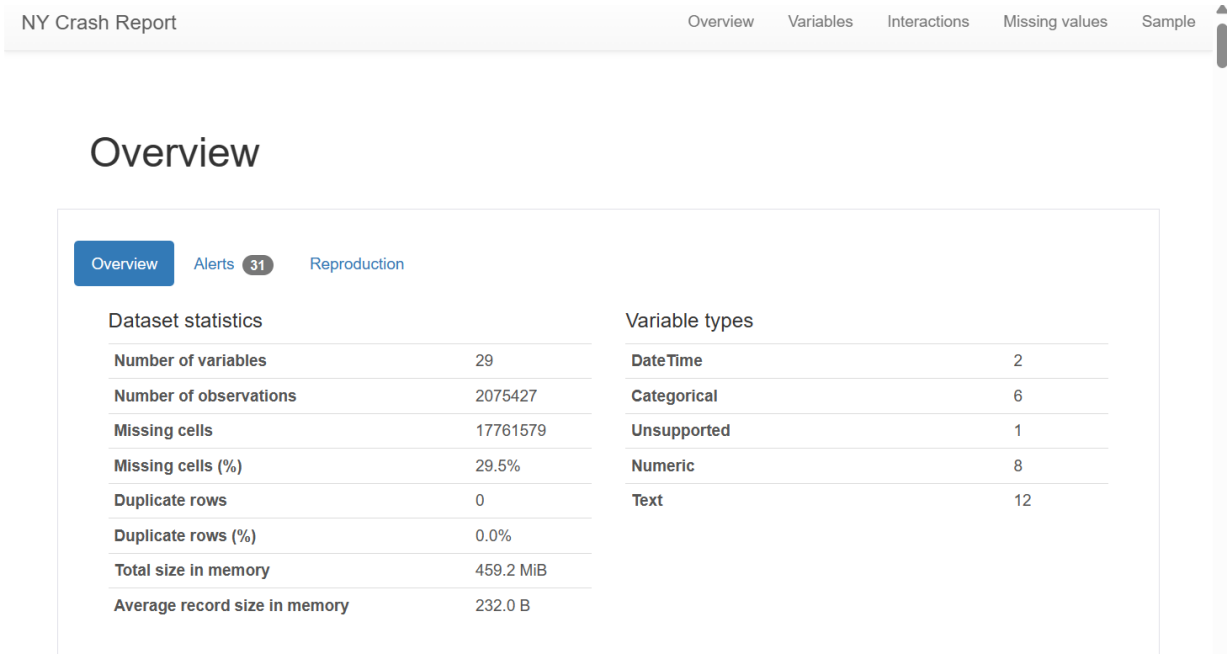


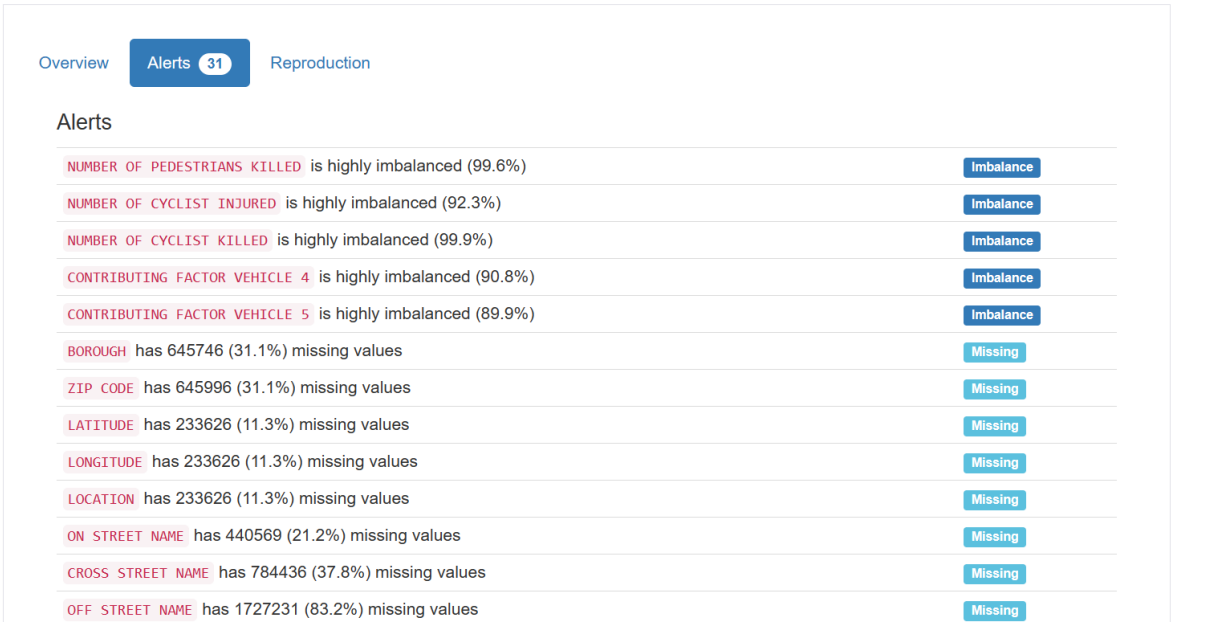
Chicago Crash Report





NY Crash Report



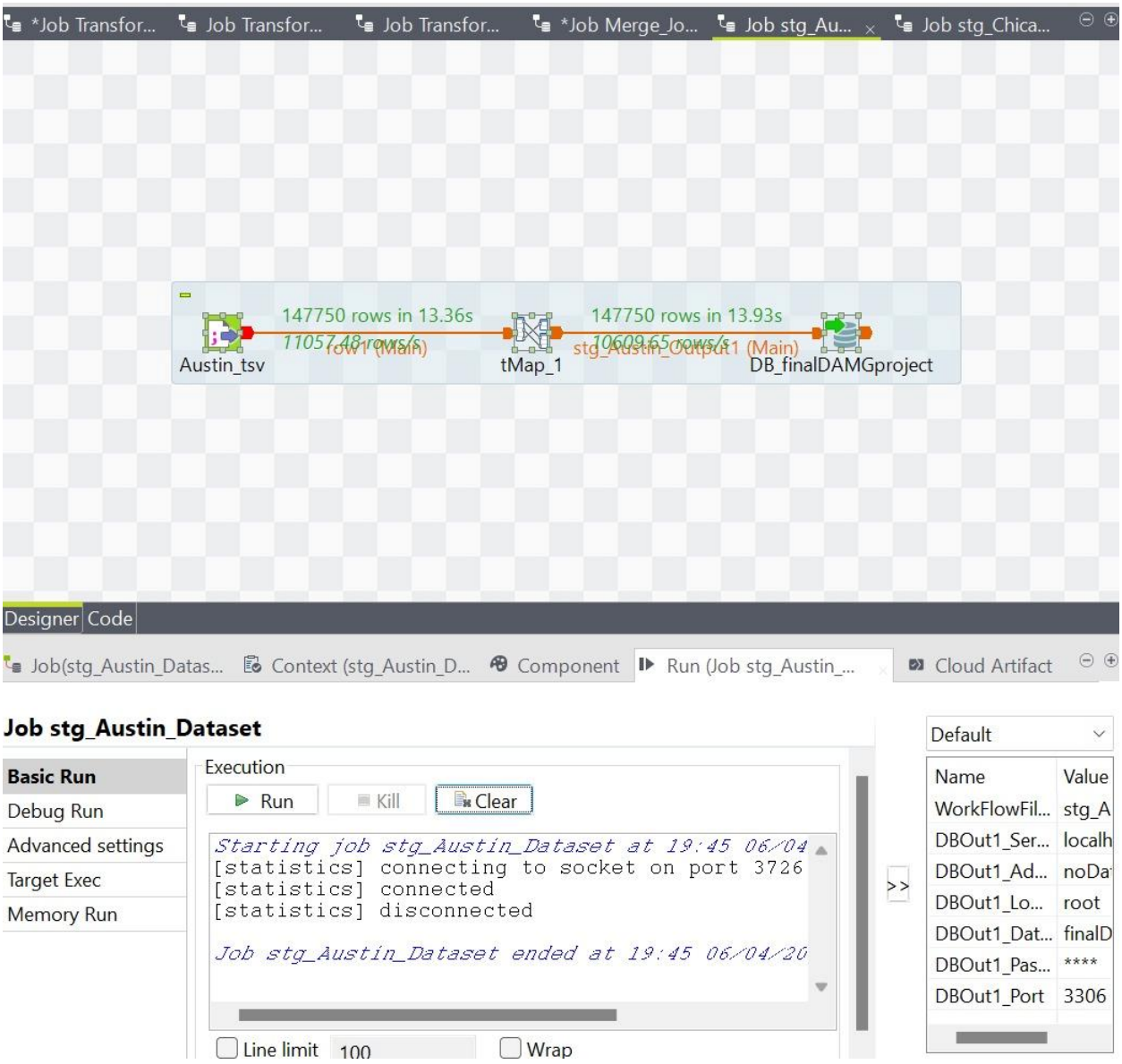


Staging

Austin

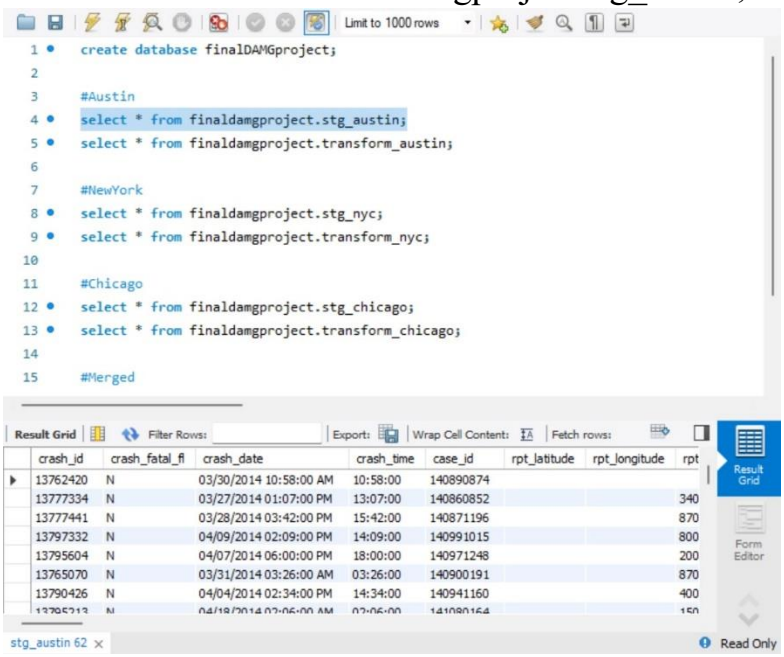
Stage Table: Stg_Austin_Dataset

Purpose: Staging Austin Dataset from tsv file

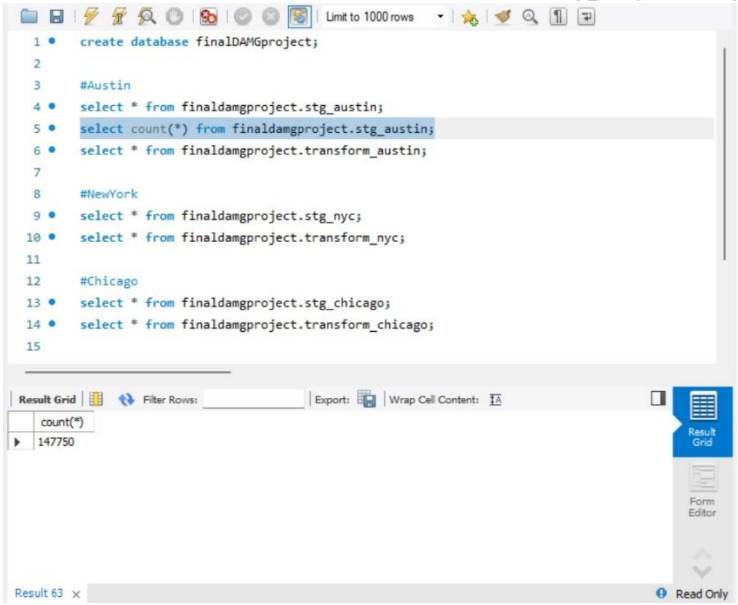


MYSQL Command

Table - select * from finaldamgproject.stg_austin;



Count - select count(*) from finaldamgproject.stg_austin



DDL commands :

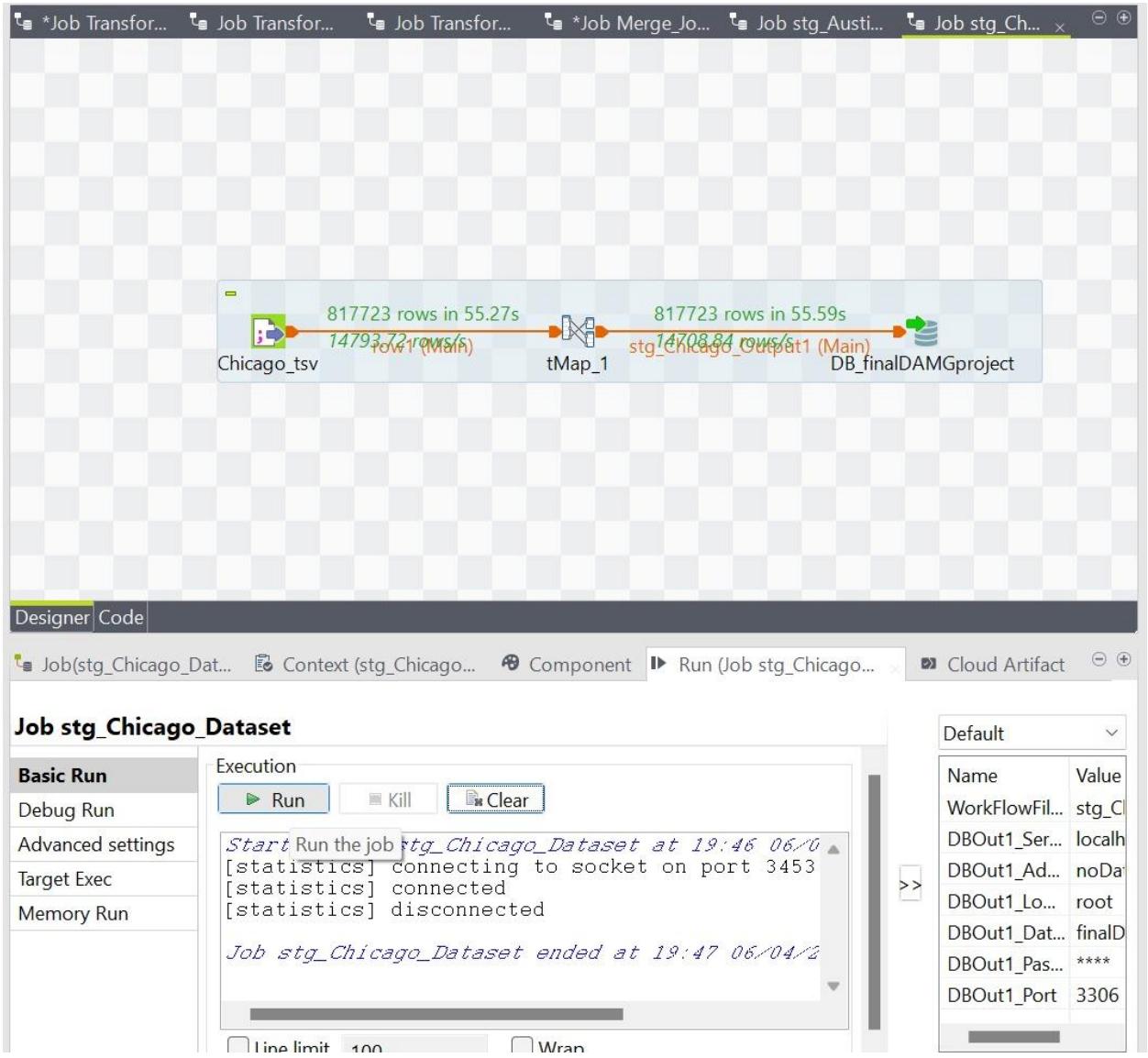
```
CREATE TABLE stg_austin (
  crash_id varchar(100) DEFAULT NULL,
  crash_fatal_fl varchar(100) DEFAULT NULL,
  crash_date varchar(100) DEFAULT NULL,
  crash_time varchar(100) DEFAULT NULL,
  case_id varchar(100) DEFAULT NULL,
  rpt_latitude varchar(100) DEFAULT NULL,
  rpt_longitude varchar(100) DEFAULT NULL,
  rpt_block_num varchar(100) DEFAULT NULL,
  rpt_street_pfx varchar(100) DEFAULT NULL,
  rpt_street_name varchar(100) DEFAULT NULL,
  rpt_street_sfx varchar(100) DEFAULT NULL,
  crash_speed_limit varchar(100) DEFAULT NULL,
  road_constr_zone_fl varchar(100) DEFAULT NULL,
  latitude varchar(100) DEFAULT NULL,
  longitude varchar(100) DEFAULT NULL,
  street_name varchar(100) DEFAULT NULL,
  street_nbr varchar(100) DEFAULT NULL,
  street_name_2 varchar(100) DEFAULT NULL,
  street_nbr_2 varchar(100) DEFAULT NULL,
  crash_sev_id int DEFAULT NULL,
  sus_serious_injry_cnt int DEFAULT NULL,
  nonincap_injry_cnt varchar(100) DEFAULT NULL,
  poss_injry_cnt varchar(100) DEFAULT NULL,
  non_injry_cnt varchar(100) DEFAULT NULL,
  unkn_injry_cnt varchar(100) DEFAULT NULL,
  tot_injry_cnt varchar(100) DEFAULT NULL,
  death_cnt int DEFAULT NULL,
  contrib_factr_p1_id varchar(100) DEFAULT NULL,
  contrib_factr_p2_id varchar(100) DEFAULT NULL,
  units_involved varchar(500) DEFAULT NULL,
  atd_mode_category_metadata varchar(6000) DEFAULT NULL,
  pedestrian_fl varchar(100) DEFAULT NULL,
  motor_vehicle_fl varchar(100) DEFAULT NULL,
```

```
  motorcycle_fl varchar(100) DEFAULT NULL,
  bicycle_fl varchar(100) DEFAULT NULL,
  other_fl varchar(100) DEFAULT NULL,
  point varchar(45) DEFAULT NULL,
  apd_confirmed_fatality varchar(100) DEFAULT NULL,
  apd_confirmed_death_count int DEFAULT NULL,
  motor_vehicle_death_count int DEFAULT NULL,
  motor_vehicle_serious_injury_count int DEFAULT NULL,
  bicycle_death_count int DEFAULT NULL,
  bicycle_serious_injury_count int DEFAULT NULL,
  pedestrian_death_count int DEFAULT NULL,
  pedestrian_serious_injury_count int DEFAULT NULL,
  motorcycle_death_count int DEFAULT NULL,
  motorcycle_serious_injury_count int DEFAULT NULL,
  other_death_count int DEFAULT NULL,
  other_serious_injury_count int DEFAULT NULL,
  onsys_fl varchar(100) DEFAULT NULL,
  private_dr_fl varchar(100) DEFAULT NULL,
  micromobility_serious_injury_count int DEFAULT NULL,
  micromobility_death_count int DEFAULT NULL,
  micromobility_fl varchar(100) DEFAULT NULL,
  DI_CreatedDate datetime DEFAULT NULL,
  DI_WorkFlowFileName varchar(100) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
COLLATE=utf8mb4_0900_ai_ci
```

Chicago

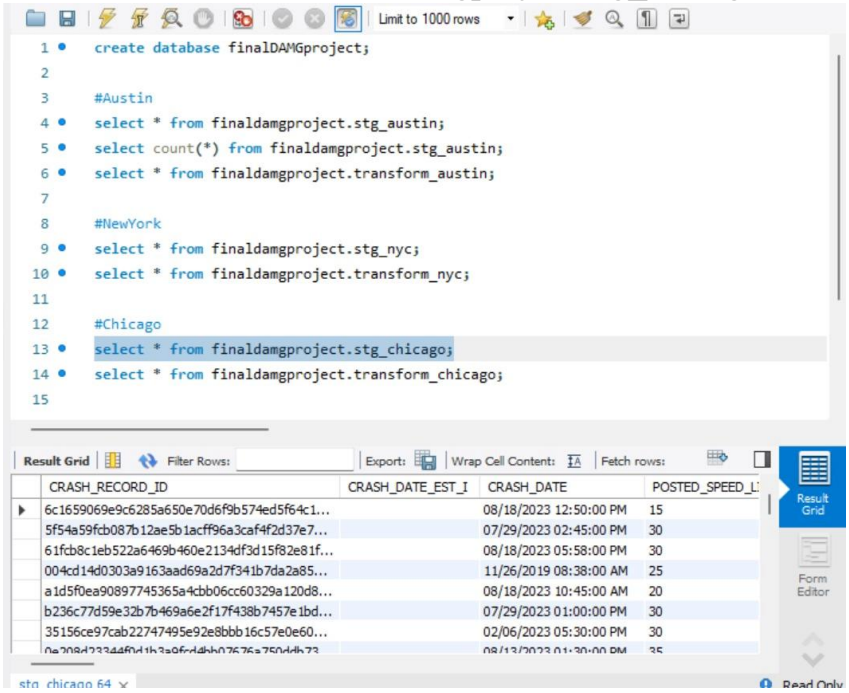
Stage Table: Stg_Chicago_Dataset

Purpose: Staging Chicago Dataset from tsv file



MYSQL Command

Table - select * from finaldamgproject.stg_chicago;

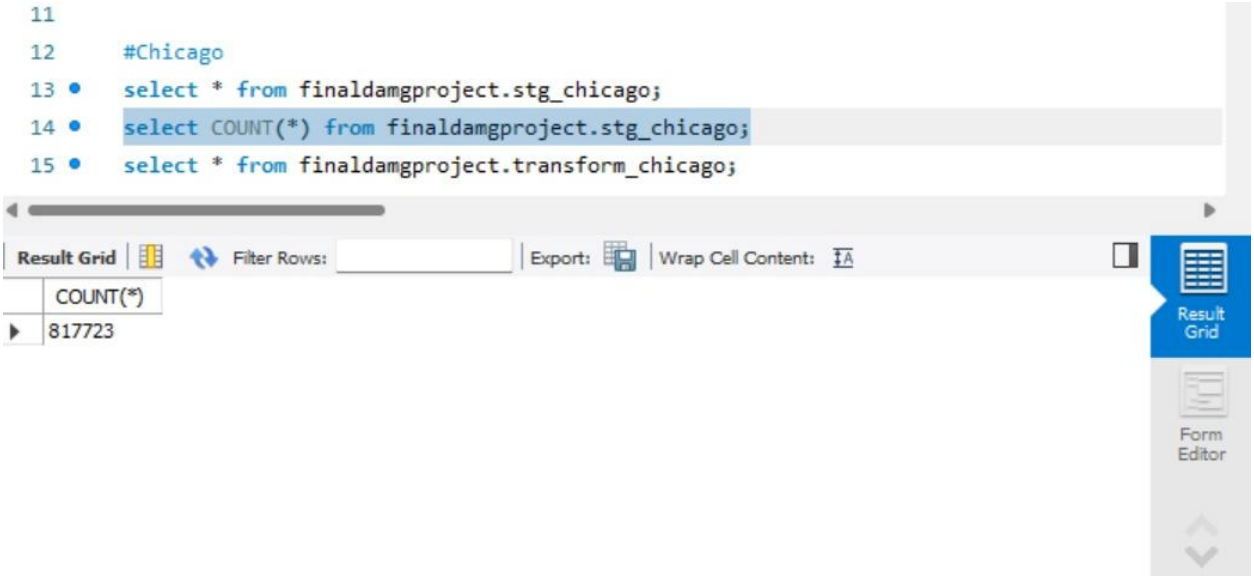


```
create database finalDAMGproject;

#Austin
select * from finaldamgproject.stg_austin;
select count(*) from finaldamgproject.stg_austin;
select * from finaldamgproject.transform_austin;

#NewYork
select * from finaldamgproject.stg_nyc;
select * from finaldamgproject.transform_nyc;

#Chicago
select * from finaldamgproject.stg_chicago;
select * from finaldamgproject.transform_chicago;
```

Result Grid

CRASH_RECORD_ID	CRASH_DATE_EST_I	CRASH_DATE	POSTED_SPEED_L
6c1659069e9c6285a650e70d6f9b574ed5f64c1...		08/18/2023 12:50:00 PM	15
5f54a59fcb087b12ae5b1acff96a3caf4f2d37e7...		07/29/2023 02:45:00 PM	30
61fcb8c1eb522a6469b460e2134df3d15f82e81f...		08/18/2023 05:58:00 PM	30
004cd14d0303a9163aad69a2d7f341b7da2a85...		11/26/2019 08:38:00 AM	25
a1d5f0ea90897745365a4cbb06cc60329a120d8...		08/18/2023 10:45:00 AM	20
b236c77d59e32b7b469a6e2f17f438b7457e1bd...		07/29/2023 01:00:00 PM	30
35156ce97cab22747495e92e8bbb16c57e0e60...		02/06/2023 05:30:00 PM	30

Count - select COUNT(*) from finaldamgproject.stg_chicago;



```
#Chicago
select * from finaldamgproject.stg_chicago;
select COUNT(*) from finaldamgproject.stg_chicago;
select * from finaldamgproject.transform_chicago;
```

Result Grid

COUNT(*)
817723

DDL commands

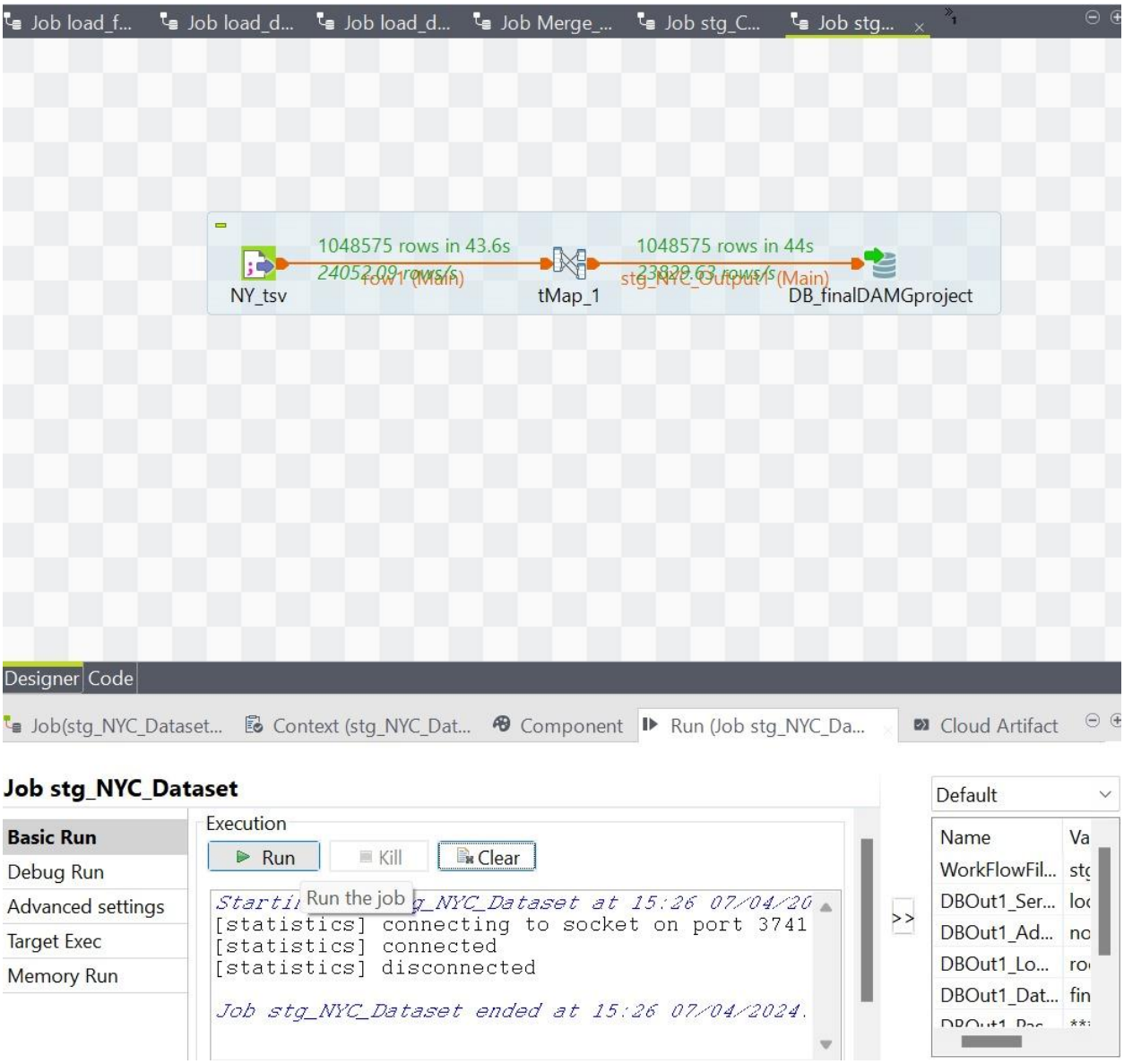
```
CREATE TABLE stg_chicago (
  CRASH_RECORD_ID varchar(500) DEFAULT NULL,
  CRASH_DATE_EST_I varchar(100) DEFAULT NULL,
  CRASH_DATE varchar(100) DEFAULT NULL,
  POSTED_SPEED_LIMIT int DEFAULT NULL,
  TRAFFIC_CONTROL_DEVICE varchar(100) DEFAULT NULL,
  DEVICE_CONDITION varchar(100) DEFAULT NULL,
  WEATHER_CONDITION varchar(100) DEFAULT NULL,
  LIGHTING_CONDITION varchar(100) DEFAULT NULL,
  FIRST_CRASH_TYPE varchar(100) DEFAULT NULL,
  TRAFFICWAY_TYPE varchar(100) DEFAULT NULL,
  LANE_CNT varchar(100) DEFAULT NULL,
  ALIGNMENT varchar(100) DEFAULT NULL,
  ROADWAY_SURFACE_COND varchar(100) DEFAULT NULL,
  ROAD_DEFECT varchar(100) DEFAULT NULL,
  REPORT_TYPE varchar(100) DEFAULT NULL,
  CRASH_TYPE varchar(100) DEFAULT NULL,
  INTERSECTION_RELATED_I varchar(100) DEFAULT NULL,
  NOT_RIGHT_OF_WAY_I varchar(100) DEFAULT NULL,
  HIT_AND_RUN_I varchar(100) DEFAULT NULL,
  DAMAGE varchar(100) DEFAULT NULL,
  DATE_POLICE_NOTIFIED varchar(100) DEFAULT NULL,
  PRIM_CONTRIBUTORY_CAUSE varchar(100) DEFAULT NULL,
  SEC_CONTRIBUTORY_CAUSE varchar(100) DEFAULT NULL,
  STREET_NO int DEFAULT NULL,
  STREET_DIRECTION varchar(100) DEFAULT NULL,
  STREET_NAME varchar(100) DEFAULT NULL,
  BEAT_OF_OCCURRENCE int DEFAULT NULL,
  PHOTOS_TAKEN_I varchar(100) DEFAULT NULL,
  STATEMENTS_TAKEN_I varchar(100) DEFAULT NULL,
  DOORING_I varchar(100) DEFAULT NULL,
  WORK_ZONE_I varchar(100) DEFAULT NULL,
  WORK_ZONE_TYPE varchar(100) DEFAULT NULL,
  WORKERS_PRESENT_I varchar(100) DEFAULT NULL,
  NUM_UNITS int DEFAULT NULL,
```

```
  MOST_SEVERE_INJURY varchar(100) DEFAULT NULL,
  INJURIES_TOTAL int DEFAULT NULL,
  INJURIES_FATAL int DEFAULT NULL,
  INJURIES_INCAPACITATING int DEFAULT NULL,
  INJURIES_NON_INCAPACITATING int DEFAULT NULL,
  INJURIES_REPORTED_NOT_EVIDENT int DEFAULT NULL,
  INJURIES_NO_INDICATION int DEFAULT NULL,
  INJURIES_UNKNOWN int DEFAULT NULL,
  CRASH_HOUR int DEFAULT NULL,
  CRASH_DAY_OF_WEEK int DEFAULT NULL,
  CRASH_MONTH int DEFAULT NULL,
  LATITUDE varchar(100) DEFAULT NULL,
  LONGITUDE varchar(100) DEFAULT NULL,
  LOCATION varchar(100) DEFAULT NULL,
  DI_CreatedDate datetime DEFAULT NULL,
  DI_WorkFlowFileName varchar(100) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
COLLATE=utf8mb4_0900_ai_ci
```

NewYork

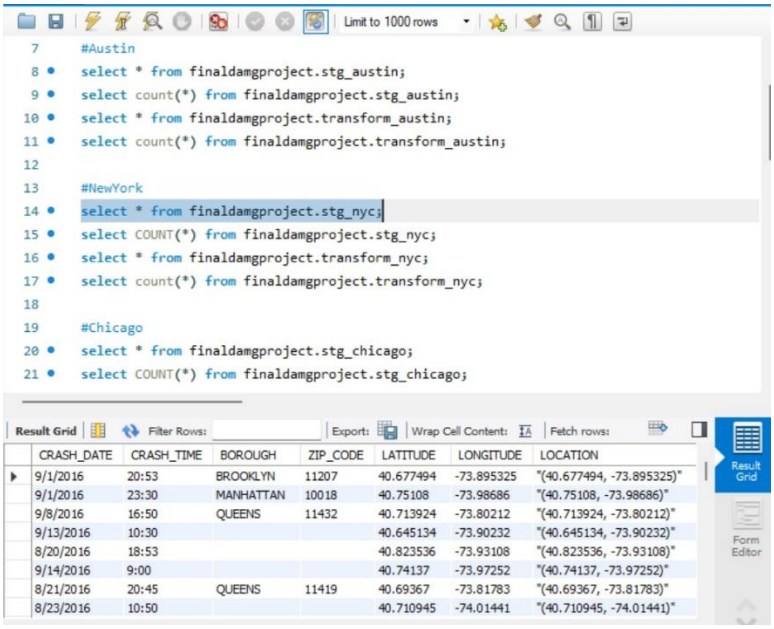
Stage Table: Stg_NYC_Dataset

Purpose: Staging New York Dataset from tsv file


Job load_f...
Job load_d...
Job load_d...
Job Merge_...
Job stg_C...
Job stg...
1048575 rows in 43.6s
24052.09 rows/s
row1 (Main)
1048575 rows in 44s
23820.63 rows/s
stg_NYC_Output1 (Main)
NY_tsv
tMap_1
DB_finalDAMGproject
Designer
Code
Job(stg_NYC_Dataset...
Context (stg_NYC_Dat...
Component
Run (Job stg_NYC_Da...
Cloud Artifact
Job stg_NYC_Dataset
Basic Run
Debug Run
Advanced settings
Target Exec
Memory Run
Execution
Run
Kill
Clear
Run the job
[statistics] connecting to socket on port 3741
[statistics] connected
[statistics] disconnected
Job stg_NYC_Dataset ended at 15:26 07/04/2024.
Default
Name
WorkFlowFil...
DBOut1_Ser...
DBOut1_Ad...
DBOut1_Lo...
DBOut1_Dat...

MYSQL Command

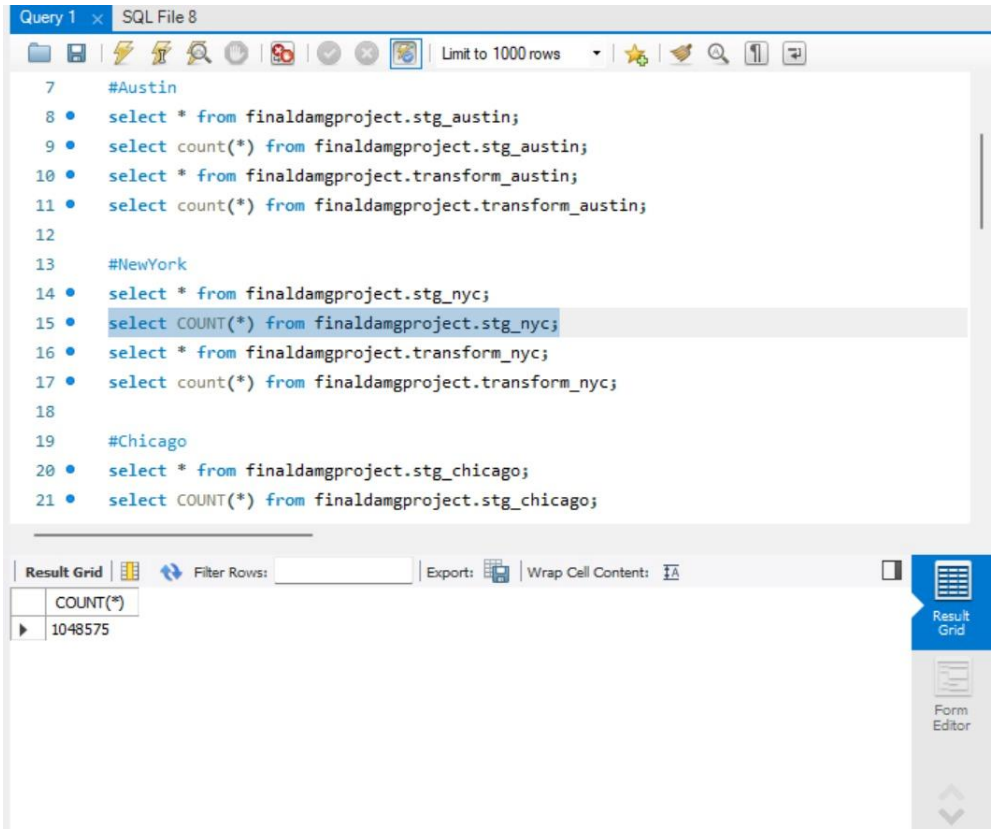
Table - select * from finaldamgproject.stg_nyc;



```
#Austin
select * from finaldamgproject.stg_austin;
select count(*) from finaldamgproject.stg_austin;
select * from finaldamgproject.transform_austin;
select count(*) from finaldamgproject.transform_austin;

#NewYork
select * from finaldamgproject.stg_nyc;
select COUNT(*) from finaldamgproject.stg_nyc;
select * from finaldamgproject.transform_nyc;
select count(*) from finaldamgproject.transform_nyc;

#Chicago
select * from finaldamgproject.stg_chicago;
select COUNT(*) from finaldamgproject.stg_chicago;
```

CRASH_DATE	CRASH_TIME	BOROUGH	ZIP_CODE	LATITUDE	LONGITUDE	LOCATION
9/1/2016	20:53	BROOKLYN	11207	40.677494	-73.895325	"(40.677494, -73.895325)"
9/1/2016	23:30	MANHATTAN	10018	40.75108	-73.98686	"(40.75108, -73.98686)"
9/8/2016	16:50	QUEENS	11432	40.713924	-73.80212	"(40.713924, -73.80212)"
9/13/2016	10:30			40.645134	-73.90232	"(40.645134, -73.90232)"
8/20/2016	18:53			40.823536	-73.93108	"(40.823536, -73.93108)"
9/14/2016	9:00			40.74137	-73.97252	"(40.74137, -73.97252)"
8/21/2016	20:45	QUEENS	11419	40.69367	-73.81783	"(40.69367, -73.81783)"
8/23/2016	10:50			40.710945	-74.01441	"(40.710945, -74.01441)"

Count - select COUNT(*) from finaldamgproject.stg_nyc;



```
#Austin
select * from finaldamgproject.stg_austin;
select count(*) from finaldamgproject.stg_austin;
select * from finaldamgproject.transform_austin;
select count(*) from finaldamgproject.transform_austin;

#NewYork
select * from finaldamgproject.stg_nyc;
select COUNT(*) from finaldamgproject.stg_nyc;
select * from finaldamgproject.transform_nyc;
select count(*) from finaldamgproject.transform_nyc;

#Chicago
select * from finaldamgproject.stg_chicago;
select COUNT(*) from finaldamgproject.stg_chicago;
```

COUNT(*)
1048575

DDL commands

```
CREATE TABLE stg_nyc (
  CRASH_DATE varchar(100) DEFAULT NULL,
  CRASH_TIME varchar(100) DEFAULT NULL,
  BOROUGH varchar(100) DEFAULT NULL,
  ZIP_CODE varchar(100) DEFAULT NULL,
  LATITUDE varchar(100) DEFAULT NULL,
  LONGITUDE varchar(100) DEFAULT NULL,
  LOCATION varchar(100) DEFAULT NULL,
  ON_STREET_NAME varchar(100) DEFAULT NULL,
  CROSS_STREET_NAME varchar(100) DEFAULT NULL,
  OFF_STREET_NAME varchar(100) DEFAULT NULL,
  NUMBER_OF_PERSONS_INJURED varchar(100) DEFAULT NULL,
  NUMBER_OF_PERSONS_KILLED int DEFAULT NULL,
  NUMBER_OF_PEDESTRIANS_INJURED int DEFAULT NULL,
  NUMBER_OF_PEDESTRIANS_KILLED int DEFAULT NULL,
  NUMBER_OF_CYCLIST_INJURED varchar(100) DEFAULT NULL,
  NUMBER_OF_CYCLIST_KILLED varchar(100) DEFAULT NULL,
  NUMBER_OF_MOTORIST_INJURED varchar(100) DEFAULT NULL,
  NUMBER_OF_MOTORIST_KILLED varchar(100) DEFAULT NULL,
  CONTRIBUTING_FACTOR_VEHICLE_1 varchar(100) DEFAULT
NULL,
  CONTRIBUTING_FACTOR_VEHICLE_2 varchar(100) DEFAULT
NULL,
  CONTRIBUTING_FACTOR_VEHICLE_3 varchar(100) DEFAULT
NULL,
  CONTRIBUTING_FACTOR_VEHICLE_4 varchar(100) DEFAULT
NULL,
  CONTRIBUTING_FACTOR_VEHICLE_5 varchar(100) DEFAULT
NULL,
  COLLISION_ID varchar(100) DEFAULT NULL,
  VEHICLE_TYPE_CODE_1 varchar(100) DEFAULT NULL,
  VEHICLE_TYPE_CODE_2 varchar(100) DEFAULT NULL,
  VEHICLE_TYPE_CODE_3 varchar(100) DEFAULT NULL,
  VEHICLE_TYPE_CODE_4 varchar(100) DEFAULT NULL,
```

```
  VEHICLE_TYPE_CODE_5 varchar(100) DEFAULT NULL,
  DI_CreatedDate datetime DEFAULT NULL,
  DI_WorkFlowFileName varchar(100) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
COLLATE=utf8mb4_0900_ai_ci
```

ContributingFactor

Stage Table: Stg_ContributingFactor_Dataset

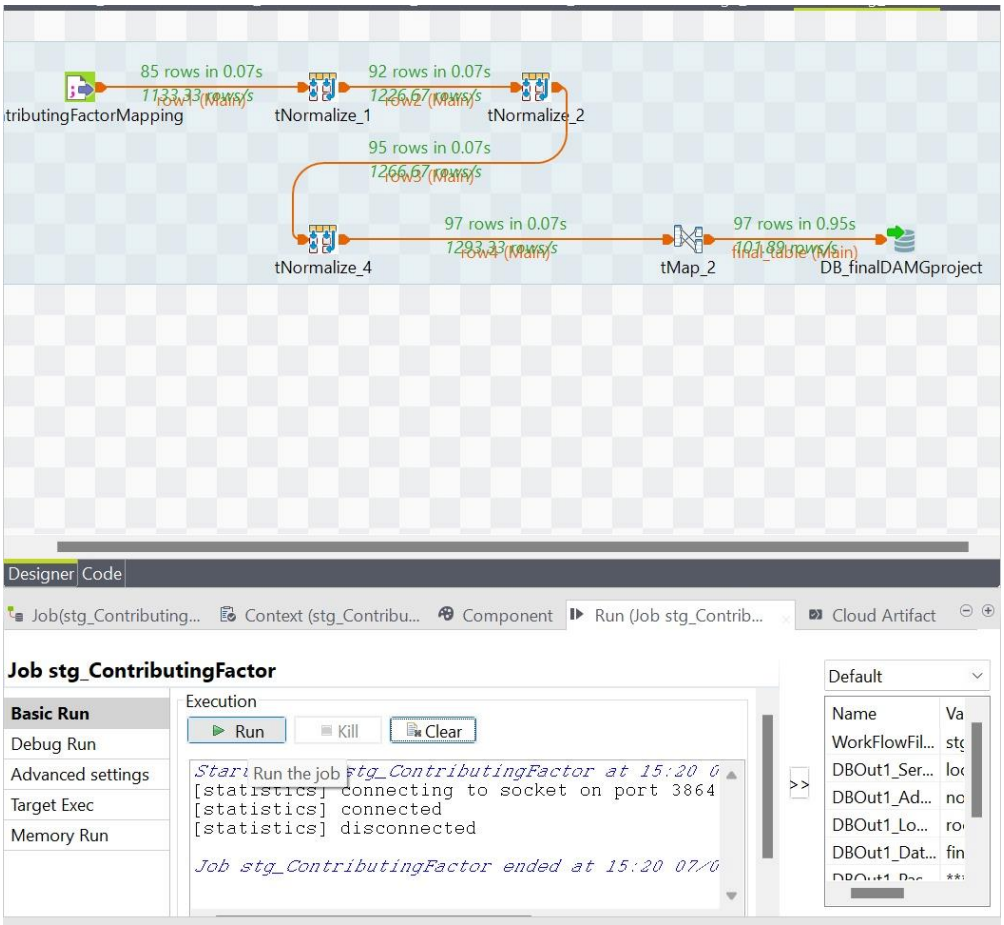
Purpose: Staging Contributing Factor Dataset from tsv file

tNormalize1: Chicago - (;)

tNormalize2: NYC - (;)

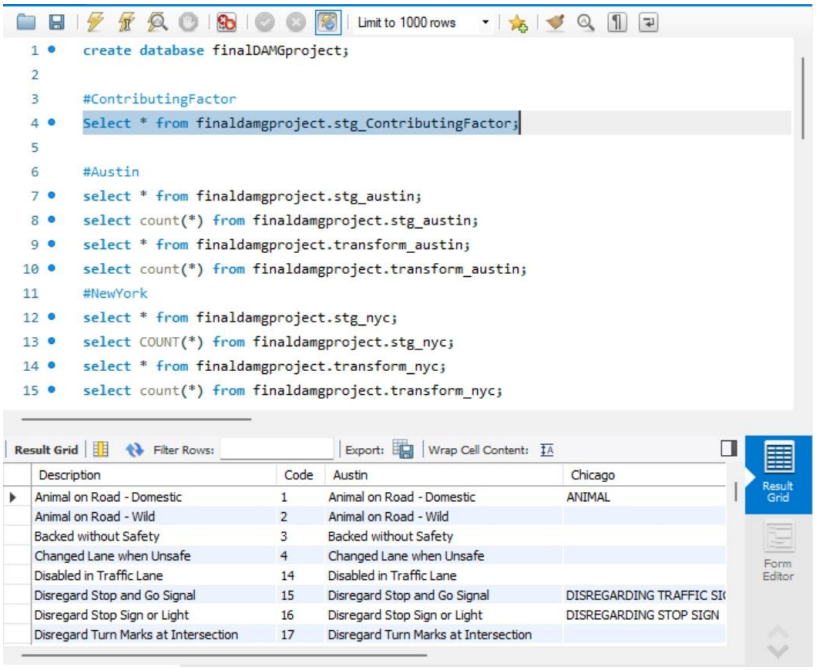
tNormalize4: NYC - (,)

tMAP >> trimming



MYSQL Command

Table - Select * from finaldamgproject.stg_ContributingFactor;



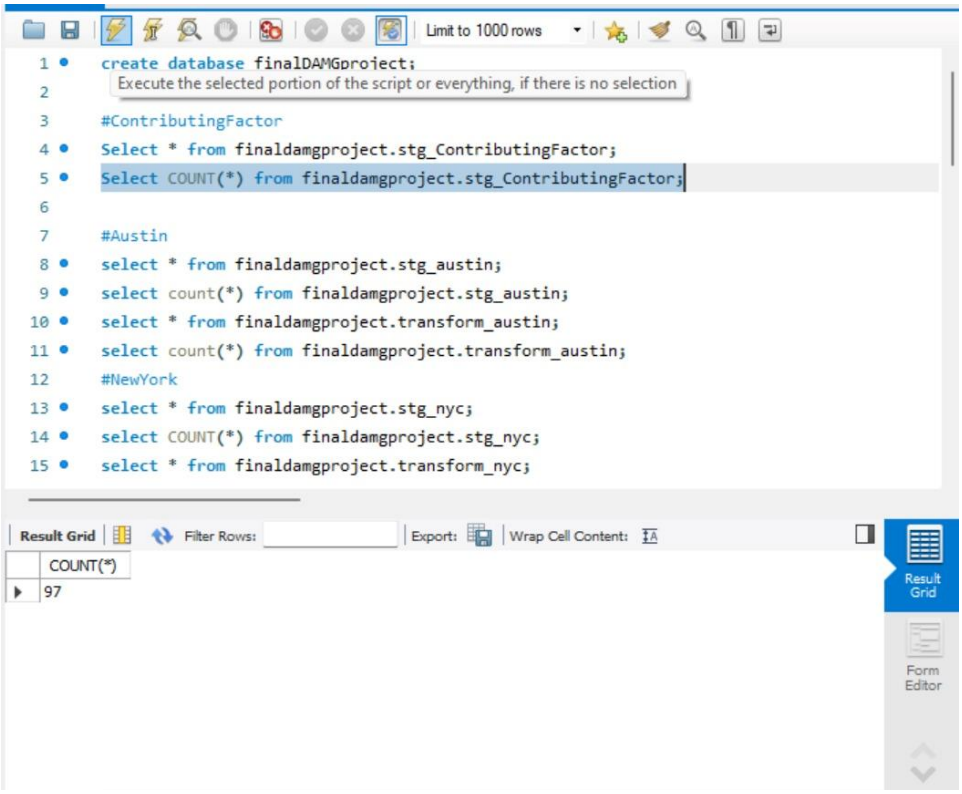
```
create database finalDAMGproject;

#ContributingFactor
Select * from finaldamgproject.stg_ContributingFactor;

#Austin
select * from finaldamgproject.stg_austin;
select count(*) from finaldamgproject.stg_austin;
select * from finaldamgproject.transform_austin;
select count(*) from finaldamgproject.transform_austin;
#NewYork
select * from finaldamgproject.stg_nyc;
select COUNT(*) from finaldamgproject.stg_nyc;
select * from finaldamgproject.transform_nyc;
select count(*) from finaldamgproject.transform_nyc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Description	Code	Austin	Chicago
Animal on Road - Domestic	1	Animal on Road - Domestic	ANIMAL
Animal on Road - Wild	2	Animal on Road - Wild	
Backed without Safety	3	Backed without Safety	
Changed Lane when Unsafe	4	Changed Lane when Unsafe	
Disabled in Traffic Lane	14	Disabled in Traffic Lane	
Disregard Stop and Go Signal	15	Disregard Stop and Go Signal	DISREGARDING TRAFFIC SI
Disregard Stop Sign or Light	16	Disregard Stop Sign or Light	DISREGARDING STOP SIGN
Disregard Turn Marks at Intersection	17	Disregard Turn Marks at Intersection	

Result Grid | Form Editor

Count - Select COUNT(*) from finaldamgproject.stg_ContributingFactor;



```
create database finalDAMGproject;
Execute the selected portion of the script or everything, if there is no selection
#ContributingFactor
Select * from finaldamgproject.stg_ContributingFactor;
Select COUNT(*) from finaldamgproject.stg_ContributingFactor;

#Austin
select * from finaldamgproject.stg_austin;
select count(*) from finaldamgproject.stg_austin;
select * from finaldamgproject.transform_austin;
select count(*) from finaldamgproject.transform_austin;
#NewYork
select * from finaldamgproject.stg_nyc;
select COUNT(*) from finaldamgproject.stg_nyc;
select * from finaldamgproject.transform_nyc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

COUNT(*)
97

Result Grid | Form Editor

DDL commands

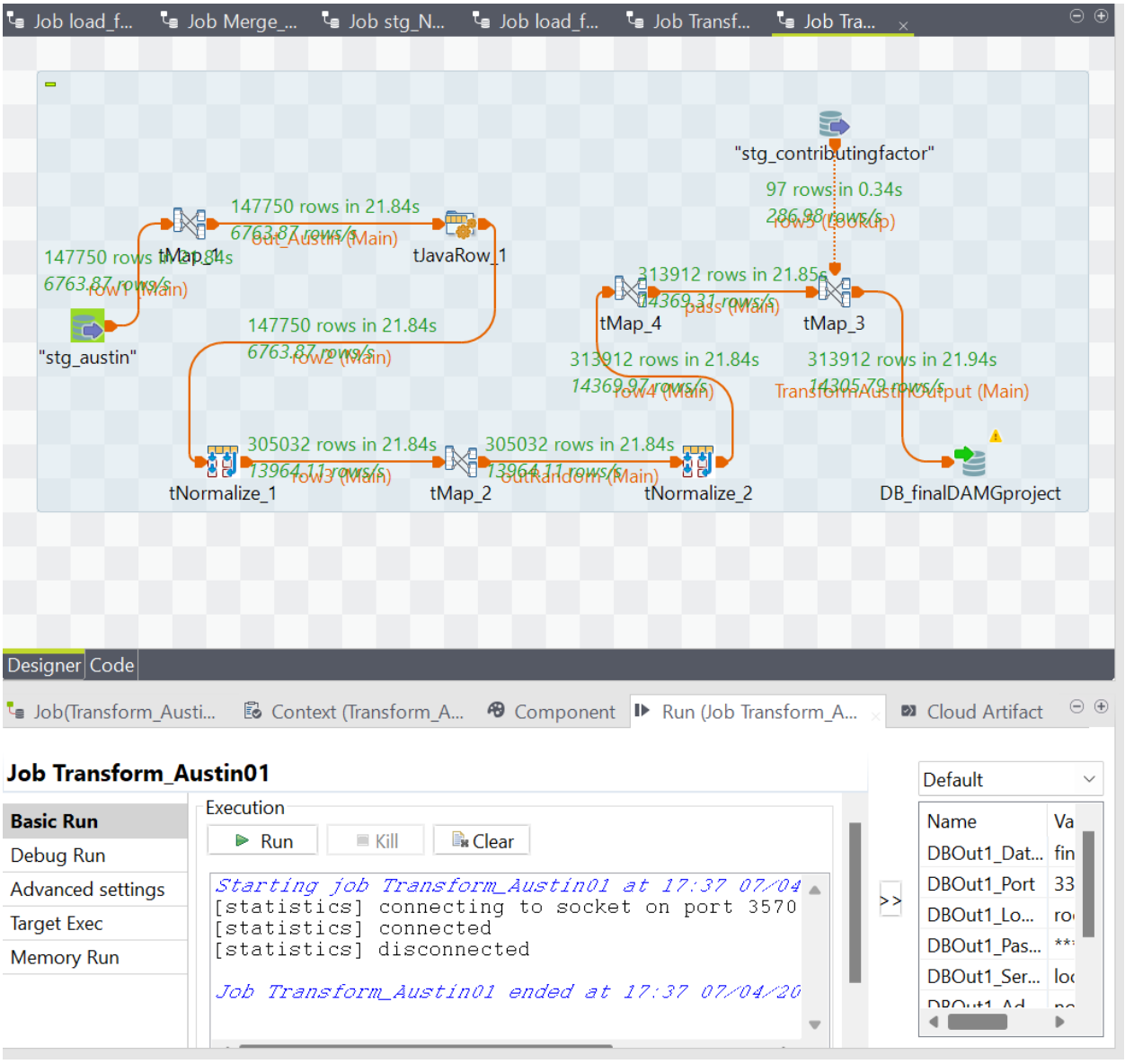
```
CREATE TABLE `stg_contributingfactor` (
  `Description` varchar(1000) DEFAULT NULL,
  `Code` int DEFAULT NULL,
  `Austin` varchar(200) DEFAULT NULL,
  `Chicago` varchar(200) DEFAULT NULL,
  `New_York` varchar(200) DEFAULT NULL,
  `DI_CreatedDate` datetime DEFAULT NULL,
  `DI_WorkFlowFileName` varchar(100) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
COLLATE=utf8mb4_0900_ai_ci
```

Transform

Austin

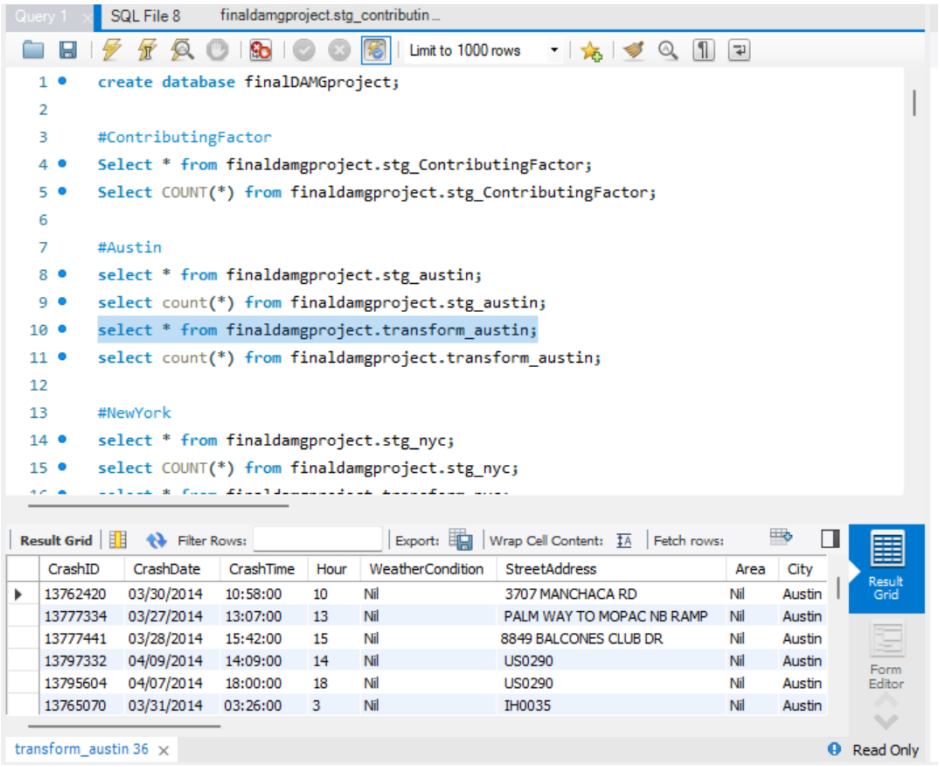
Transform Table: Transforming_Austin

Purpose
tJavaRow: Calculate total vehicle count
tNormalize: normalize Vehicle Type
tNormalize: Factor Code
tMAP: joining Austin to contributing factor to get Description

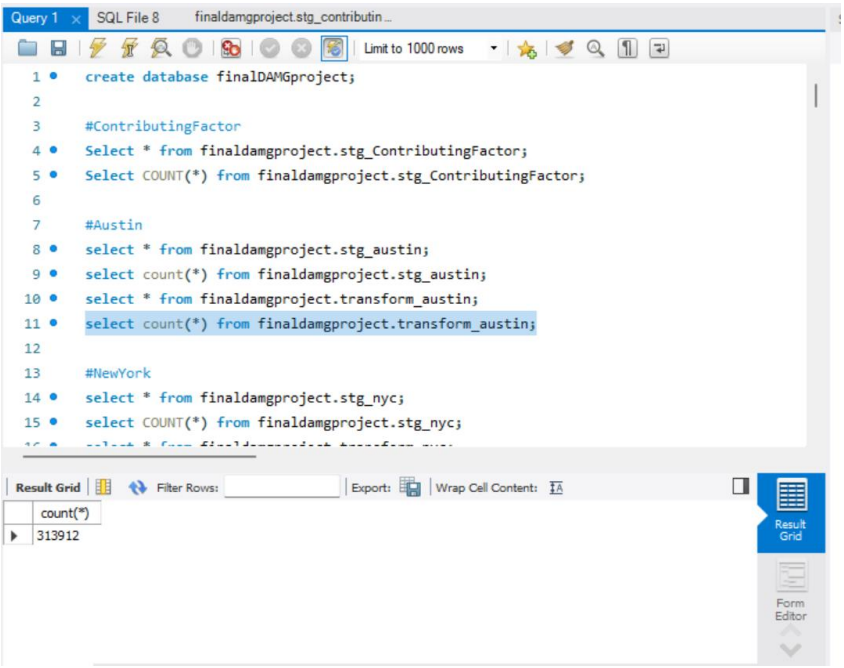

Job load_f...
Job Merge_...
Job stg_N...
Job load_f...
Job Transf...
Job Tra...
"stg_contributingfactor"
97 rows in 0.34s
286.98 rows/s
row5 (Lookup)
147750 rows in 21.84s
6763.87 rows/s
out_Austin (Main)
tMap_1
tJavaRow_1
147750 rows in 21.84s
6763.87 rows/s
row1 (Main)
"stg_austin"
313912 rows in 21.85s
14369.31 rows/s
pass (Main)
tMap_4
tMap_3
147750 rows in 21.84s
6763.87 rows/s
row2 (Main)
313912 rows in 21.84s
14369.97 rows/s
row4 (Main)
313912 rows in 21.94s
14305.79 rows/s
TransformAustinOutput (Main)
305032 rows in 21.84s
13964.11 rows/s
row3 (Main)
305032 rows in 21.84s
13964.11 rows/s
outRandom (Main)
tNormalize_1
tMap_2
tNormalize_2
DB_finalDAMGproject
Designer
Code
Job(Transform_Austi...
Context (Transform_A...
Component
Run (Job Transform_A...
Cloud Artifact
Job Transform_Austin01
Basic Run
Debug Run
Advanced settings
Target Exec
Memory Run
Execution
Run
Kill
Clear
Starting job Transform_Austin01 at 17:37 07/04
[statistics] connecting to socket on port 3570
[statistics] connected
[statistics] disconnected
Job Transform_Austin01 ended at 17:37 07/04/20
Default
Name
Va
DBOut1_Dat...
fin
DBOut1_Port
33
DBOut1_Lo...
DBOut1_Pas...
DBOut1_Ser...
loc

MYSQL Command

Table



Count



DDL commands

```
CREATE TABLE `transform_austin` (
  `CrashID` varchar(500) DEFAULT NULL,
  `CrashDate` varchar(20) DEFAULT NULL,
  `CrashTime` varchar(10) DEFAULT NULL,
  `Hour` int DEFAULT NULL,
  `WeatherCondition` varchar(100) DEFAULT NULL,
  `StreetAddress` varchar(500) DEFAULT NULL,
  `Area` varchar(20) DEFAULT NULL,
  `City` varchar(20) DEFAULT NULL,
  `ZipCode` varchar(10) DEFAULT NULL,
  `Latitude` varchar(20) DEFAULT NULL,
  `Longitude` varchar(20) DEFAULT NULL,
  `SourceName` varchar(50) DEFAULT NULL,
  `MotoristKilled` int DEFAULT NULL,
  `MotoristInjured` int DEFAULT NULL,
  `PedastriansKilled` int DEFAULT NULL,
  `PedastriansInjured` int DEFAULT NULL,
  `CyclistKilled` int DEFAULT NULL,
  `CyclistInjured` int DEFAULT NULL,
  `TotalKilled` int DEFAULT NULL,
  `TotalInjured` int DEFAULT NULL,
  `VehicleType` varchar(250) DEFAULT NULL,
  `VehicleCount` int DEFAULT NULL,
  `FactorCode` int DEFAULT NULL,
  `FactorDescription` varchar(1000) DEFAULT NULL,
  `DI_CreatedDate` datetime DEFAULT NULL,
  `DI_WorkFlowFileName` varchar(100) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
COLLATE=utf8mb4_0900_ai_ci
```

Chicago

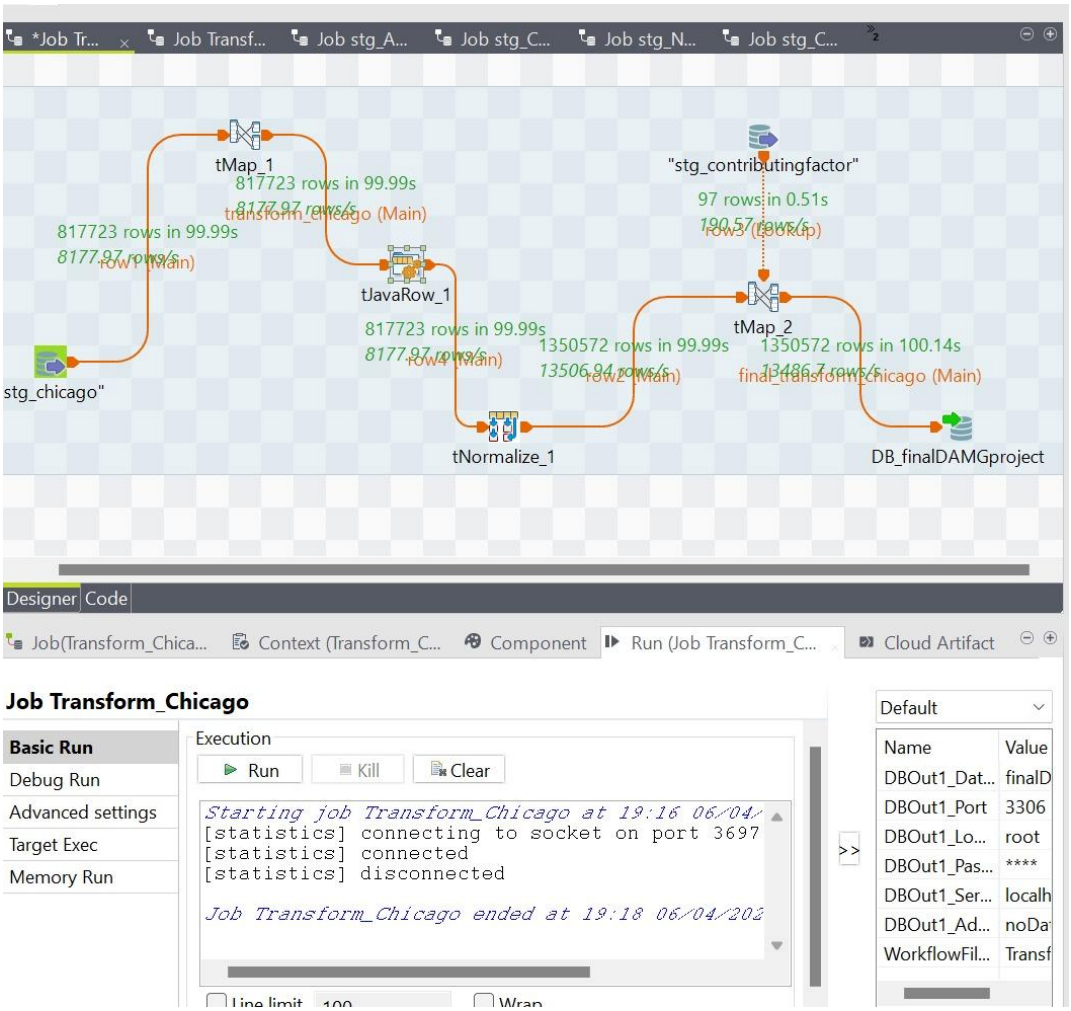
Transform Table: Transforming_Chicago

Purpose

tJavaRow: Denormalize factor description

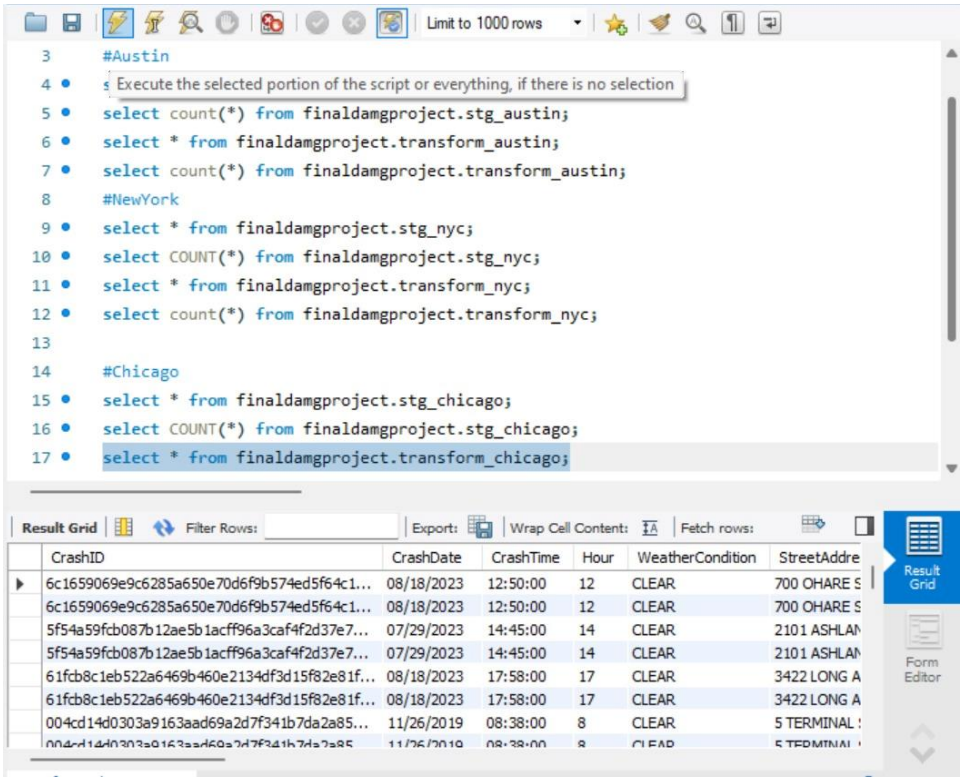
tNormalize: Factor Description

tMAP: joining Chicago to contributing factor to get Code

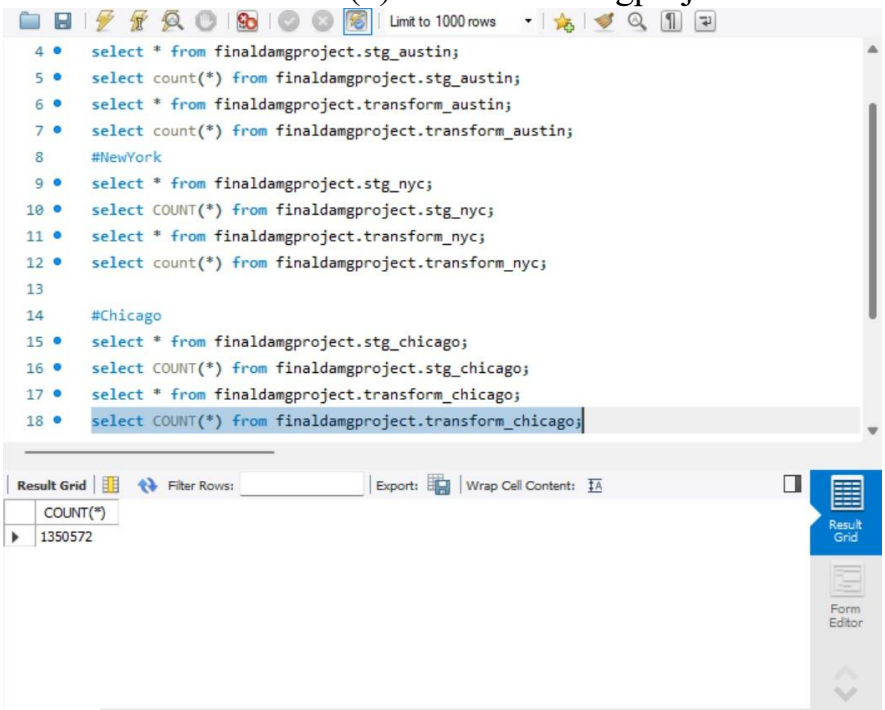


MYSQL Command

Table - select * from finaldamgproject.transform_chicago;



Count - select COUNT(*) from finaldamgproject.transform_chicago;



DDL commands

```
CREATE TABLE transform_chicago (
  CrashID varchar(500) DEFAULT NULL,
  CrashDate varchar(20) DEFAULT NULL,
  CrashTime varchar(10) DEFAULT NULL,
  Hour int DEFAULT NULL,
  WeatherCondition varchar(100) DEFAULT NULL,
  StreetAddress varchar(500) DEFAULT NULL,
  Area varchar(20) DEFAULT NULL,
  City varchar(20) DEFAULT NULL,
  ZipCode varchar(10) DEFAULT NULL,
  Latitude varchar(20) DEFAULT NULL,
  Longitude varchar(20) DEFAULT NULL,
  SourceName varchar(50) DEFAULT NULL,
  MotoristKilled int DEFAULT NULL,
  MotoristInjured int DEFAULT NULL,
  PedastriansKilled int DEFAULT NULL,
  PedastriansInjured int DEFAULT NULL,
  CyclistKilled int DEFAULT NULL,
  CyclistInjured int DEFAULT NULL,
  TotalKilled int DEFAULT NULL,
  TotalInjured int DEFAULT NULL,
  VehicleType varchar(250) DEFAULT NULL,
  VehicleCount int DEFAULT NULL,
  FactorCode int DEFAULT NULL,
  FactorDescription varchar(1000) DEFAULT NULL,
  DI_CreatedDate datetime DEFAULT NULL,
  DI_WorkFlowFileName varchar(100) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
COLLATE=utf8mb4_0900_ai_ci
```

NewYork

Transform Table: Transforming_NYC

Purpose

tFilterRow: 2 Rows Rejecting

Collision_ID : empty – 2 rows

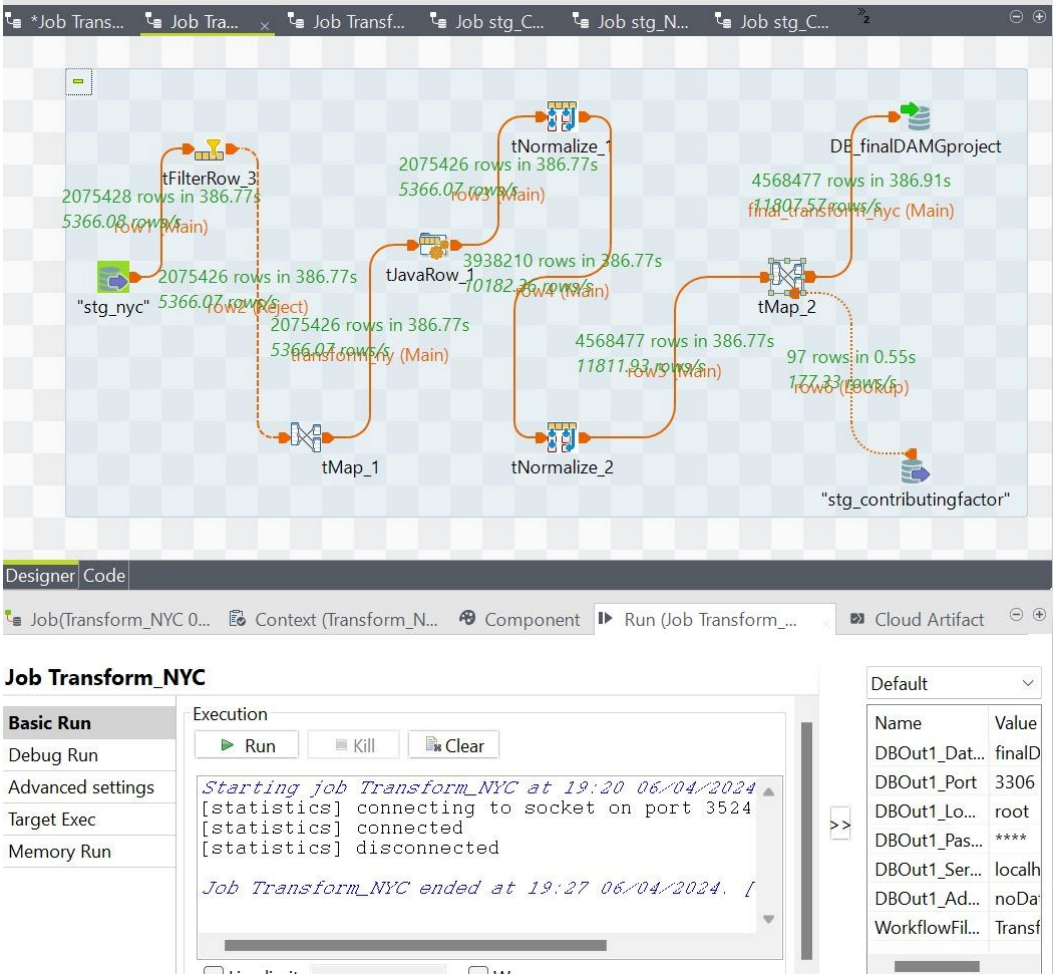
Number of Cyclist – Passenger Vehicle – Dirty info

tJavaRow: Denormalize Vehicle Type and Factor Description

tNormalize: normalize Vehicle Type

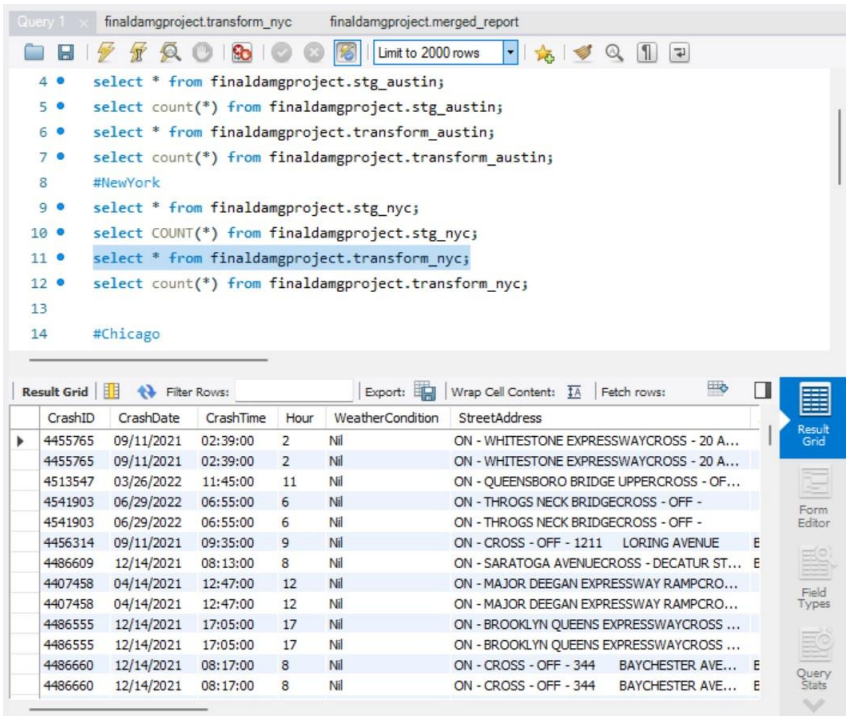
tNormalize: Factor Factor Description

tMAP: joining NYC to contributing factor

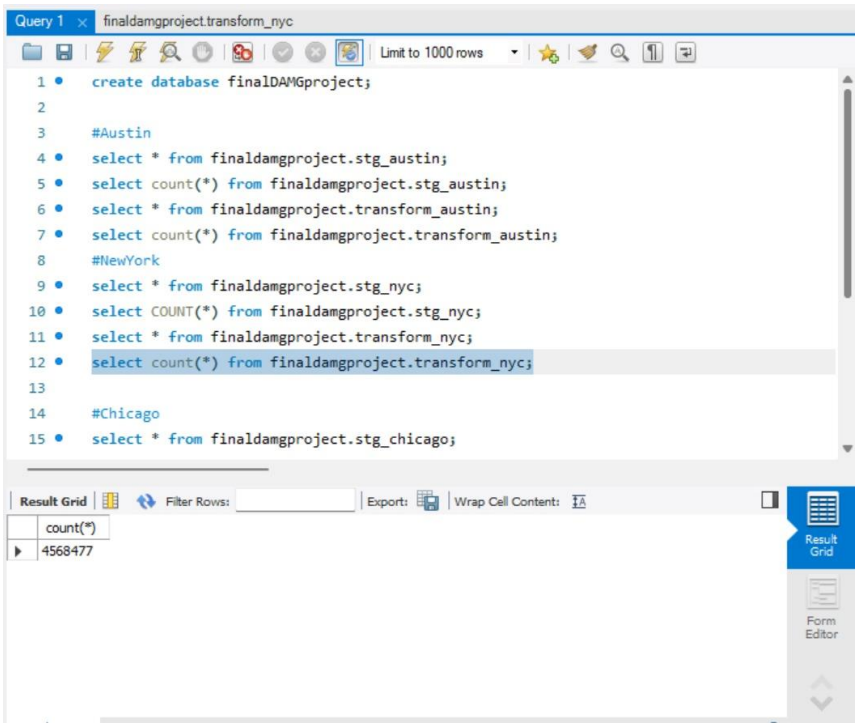


MYSQL Command

Table - select * from finaldamgproject.transform_nyc;



Count - select count(*) from finaldamgproject.transform_nyc;



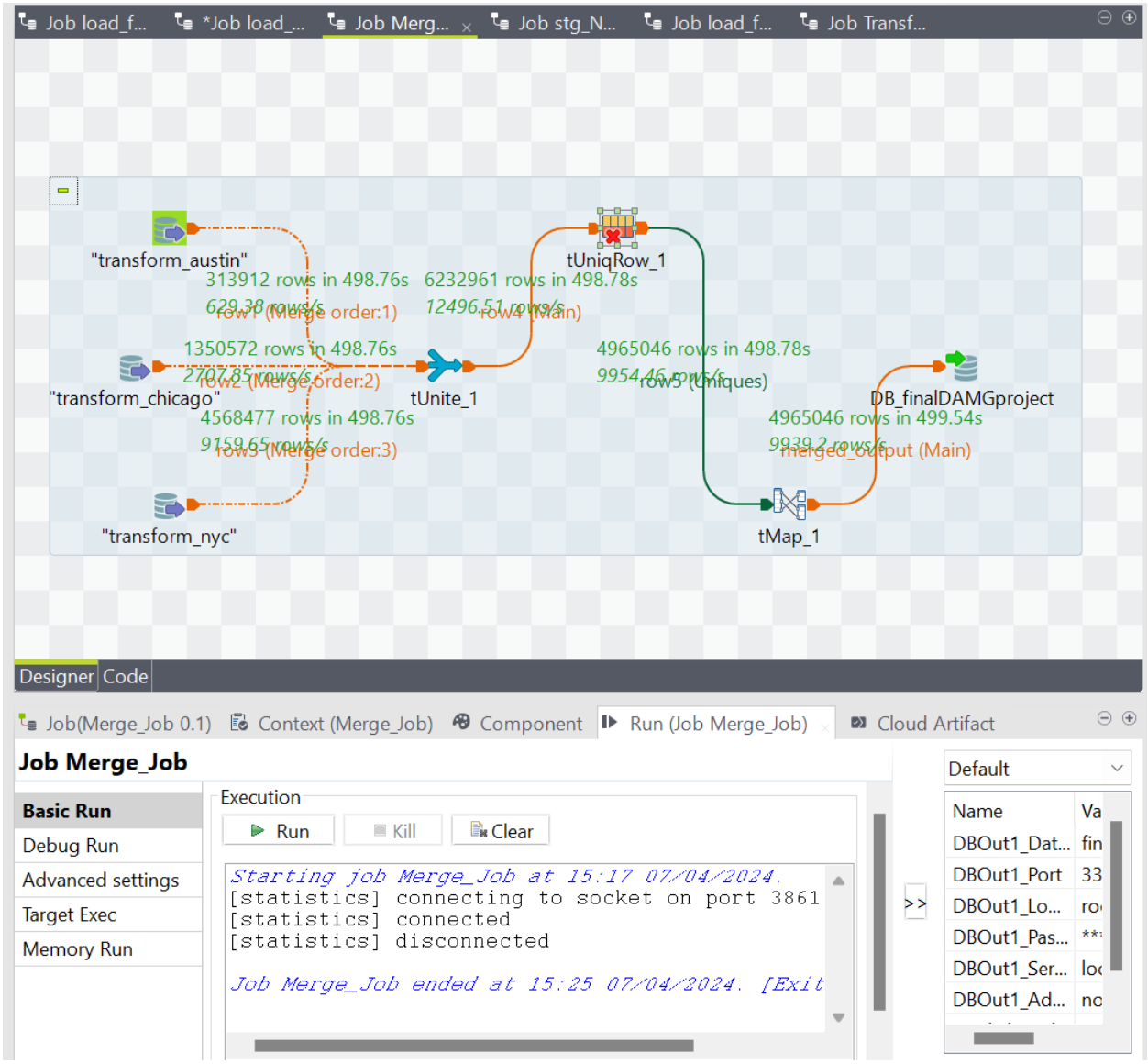
DDL commands

```
CREATE TABLE transform_nyc (
  CrashID varchar(500) DEFAULT NULL,
  CrashDate varchar(20) DEFAULT NULL,
  CrashTime varchar(10) DEFAULT NULL,
  Hour int DEFAULT NULL,
  WeatherCondition varchar(100) DEFAULT NULL,
  StreetAddress varchar(500) DEFAULT NULL,
  Area varchar(20) DEFAULT NULL,
  City varchar(20) DEFAULT NULL,
  ZipCode varchar(10) DEFAULT NULL,
  Latitude varchar(20) DEFAULT NULL,
  Longitude varchar(20) DEFAULT NULL,
  SourceName varchar(50) DEFAULT NULL,
  MotoristKilled int DEFAULT NULL,
  MotoristInjured int DEFAULT NULL,
  PedastriansKilled int DEFAULT NULL,
  PedastriansInjured int DEFAULT NULL,
  CyclistKilled int DEFAULT NULL,
  CyclistInjured int DEFAULT NULL,
  TotalKilled int DEFAULT NULL,
  TotalInjured int DEFAULT NULL,
  VehicleType varchar(250) DEFAULT NULL,
  VehicleCount int DEFAULT NULL,
  FactorCode int DEFAULT NULL,
  FactorDescription varchar(1000) DEFAULT NULL,
  DI_CreatedDate datetime DEFAULT NULL,
  DI_WorkFlowFileName varchar(100) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
COLLATE=utf8mb4_0900_ai_ci
```

Merge

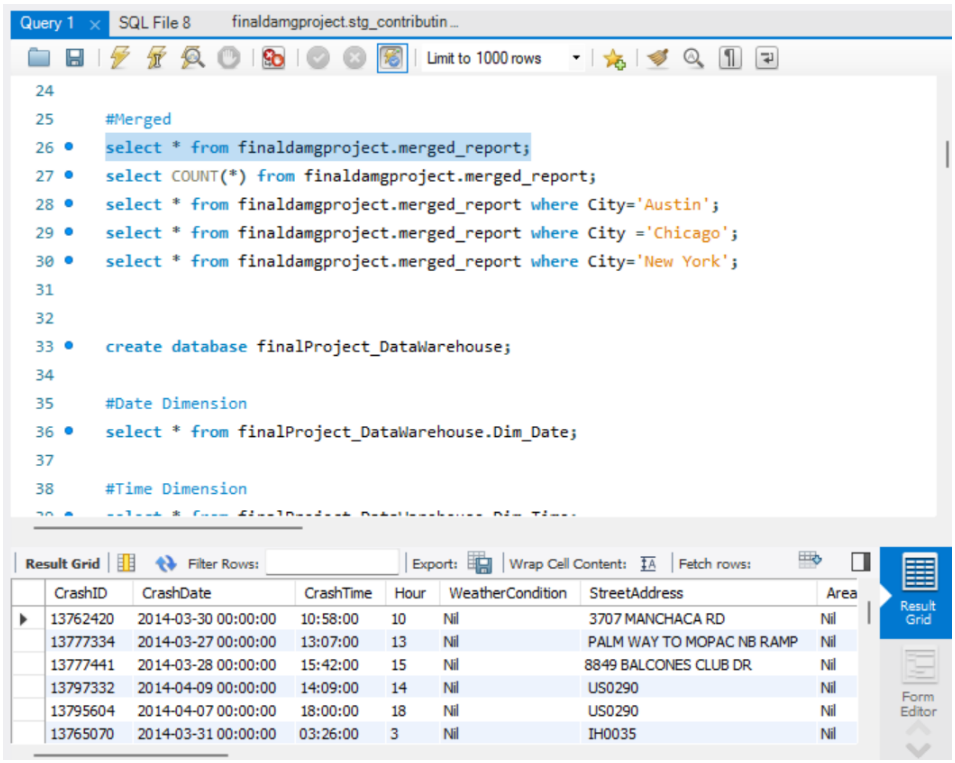
Merging Table: Merge_Job

Purpose: Uniting transform_austin, transform_chicago, transform_nyc

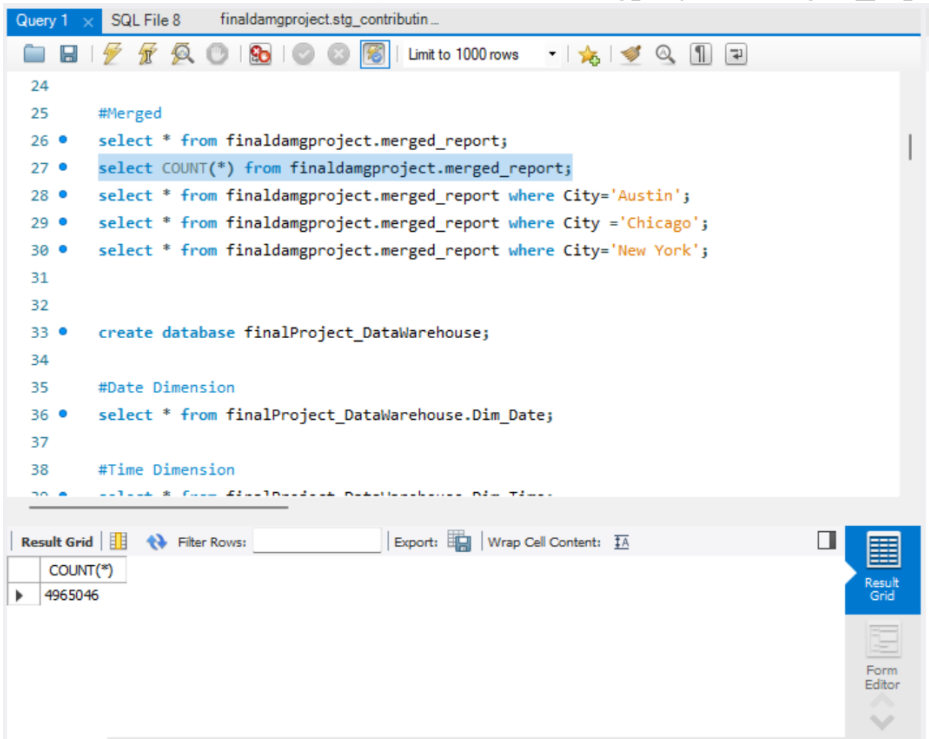


MYSQL Command

Table : select * from finaldamgproject.merged_report;



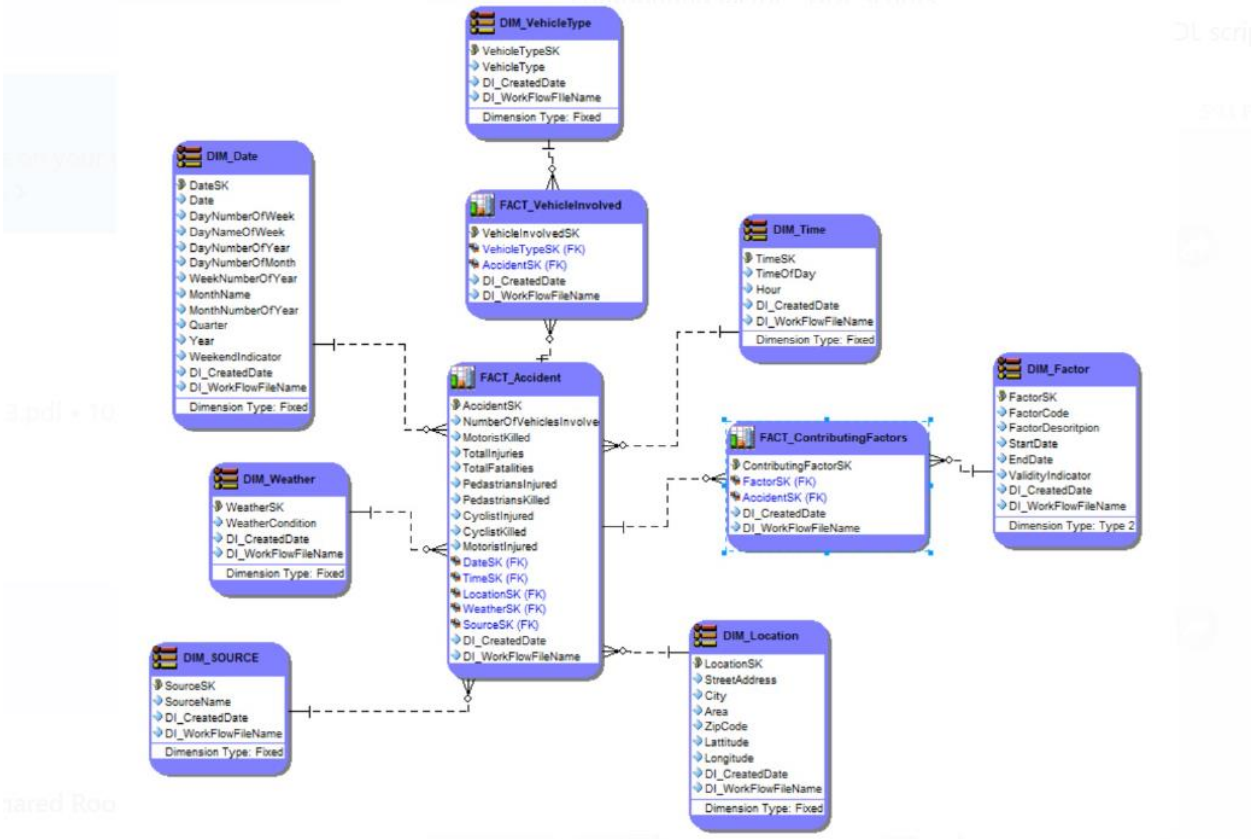
Count : select COUNT(*) from finaldamgproject.merged_report;



DDL commands

```
CREATE TABLE merged_report (
  CrashID varchar(500) DEFAULT NULL,
  CrashDate varchar(20) DEFAULT NULL,
  CrashTime varchar(10) DEFAULT NULL,
  Hour int DEFAULT NULL,
  WeatherCondition varchar(100) DEFAULT NULL,
  StreetAddress varchar(500) DEFAULT NULL,
  Area varchar(20) DEFAULT NULL,
  City varchar(20) DEFAULT NULL,
  ZipCode varchar(10) DEFAULT NULL,
  Latitude varchar(20) DEFAULT NULL,
  Longitude varchar(20) DEFAULT NULL,
  SourceName varchar(50) DEFAULT NULL,
  MotoristKilled int DEFAULT NULL,
  MotoristInjured int DEFAULT NULL,
  PedastriansKilled int DEFAULT NULL,
  PedastriansInjured int DEFAULT NULL,
  CyclistKilled int DEFAULT NULL,
  CyclistInjured int DEFAULT NULL,
  TotalKilled int DEFAULT NULL,
  TotalInjured int DEFAULT NULL,
  VehicleType varchar(250) DEFAULT NULL,
  VehicleCount int DEFAULT NULL,
  FactorCode int DEFAULT NULL,
  FactorDescription varchar(1000) DEFAULT NULL,
  DI_CreatedDate datetime DEFAULT NULL,
  DI_WorkFlowFileName varchar(100) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
COLLATE=utf8mb4_0900_ai_ci
```

Dimensional Model



Dimension Tables:

- DIM_Date: Contains attributes like DateSK, Date, DayNumberOfWeek, DayNameOfWeek, DayNumberOfYear, MonthName, Quarter, Year, WeekendIndicator, DI_CreatedDate, DI_WorkFlowFileName.
- DIM_Time: Includes TimeSK, TimeOfDay, Hour, DI_CreatedDate, DI_WorkFlowFileName.
- DIM_Weather: Holds WeatherSK, WeatherCondition, DI_CreatedDate, DI_WorkFlowFileName.
- DIM_Source: Contains SourceSK, SourceName, DI_CreatedDate, DI_WorkFlowFileName.

- DIM_Location: Contains attributes like LocationSK, StreetAddress, Area, City, ZipCode, Latitude, Longitude, DI_CreatedDate, DI_WorkFlowFileName.
- DIM_VehicleType: Comprises VehicleTypeSK, VehicleType, DI_CreatedDate, DI_WorkFlowFileName.
- DIM_Factor - SCD Type 2 dimension - Contains FactorSK, FactorCode, FactorDescription, StartDate, EndDate, ValidityIndicator, DI_CreatedDate, DI_WorkFlowFileName.

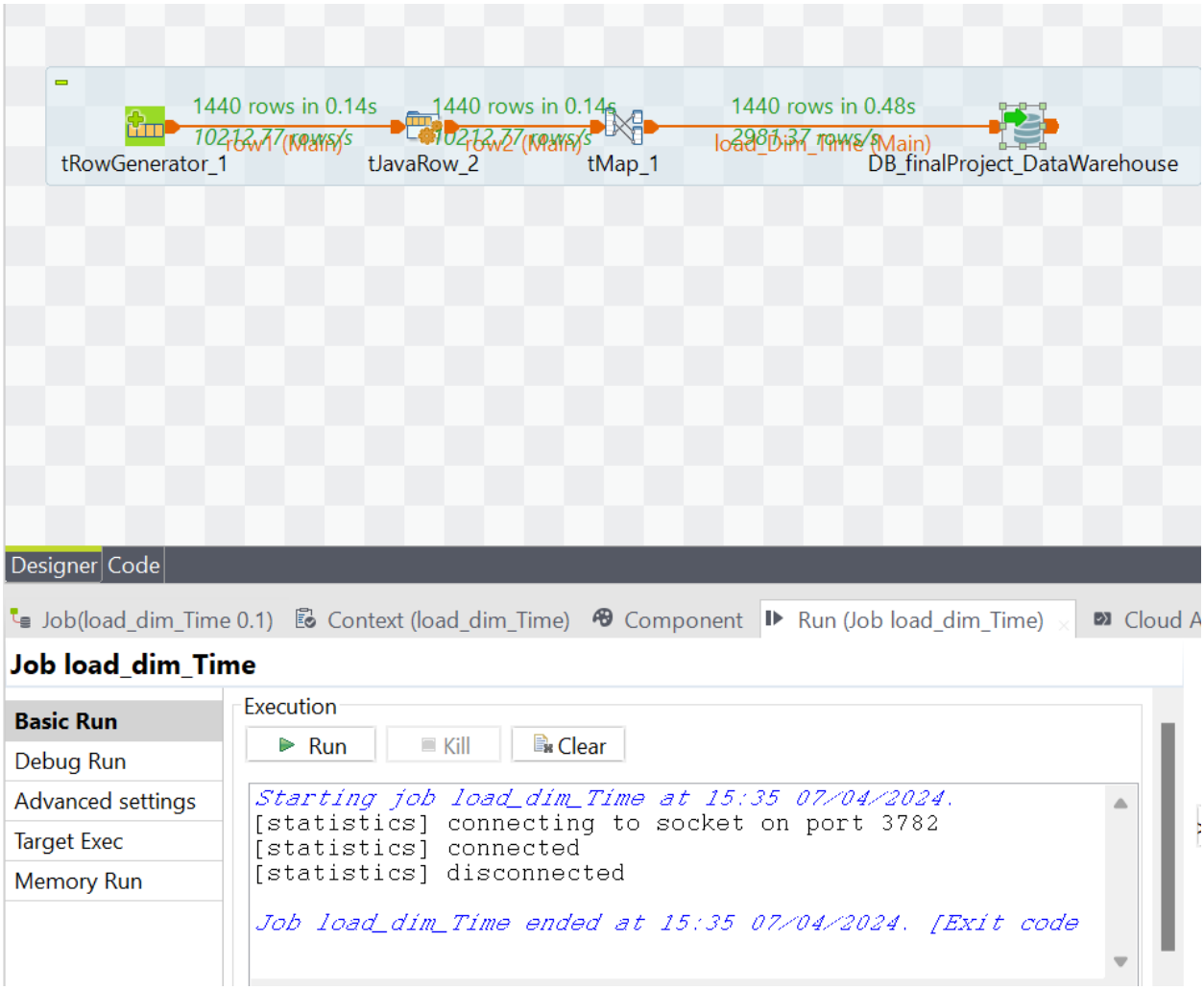
Note: Choosing DIM_Factor as an SCD Type 2 is strategic because the nature of contributing factors to accidents may evolve over time, and the impact of these changes on accident analysis is significant. For example:
A contributing factor might be redefined or split into two separate factors, and analysts would need to distinguish accidents occurring under the old definition from those under the new one.
If a factor is no longer relevant, it can be ended without losing the historical data associated with it.
New factors might emerge, which would need to be tracked separately from the existing ones.
SCD Type 2 ensures that historical data integrity is maintained, and trends and analyses remain accurate even as the descriptions and relevancy of factors change over time. It allows the reporting to accurately reflect the context and details of contributing factors as they were at the time of each accident.

Fact Tables:

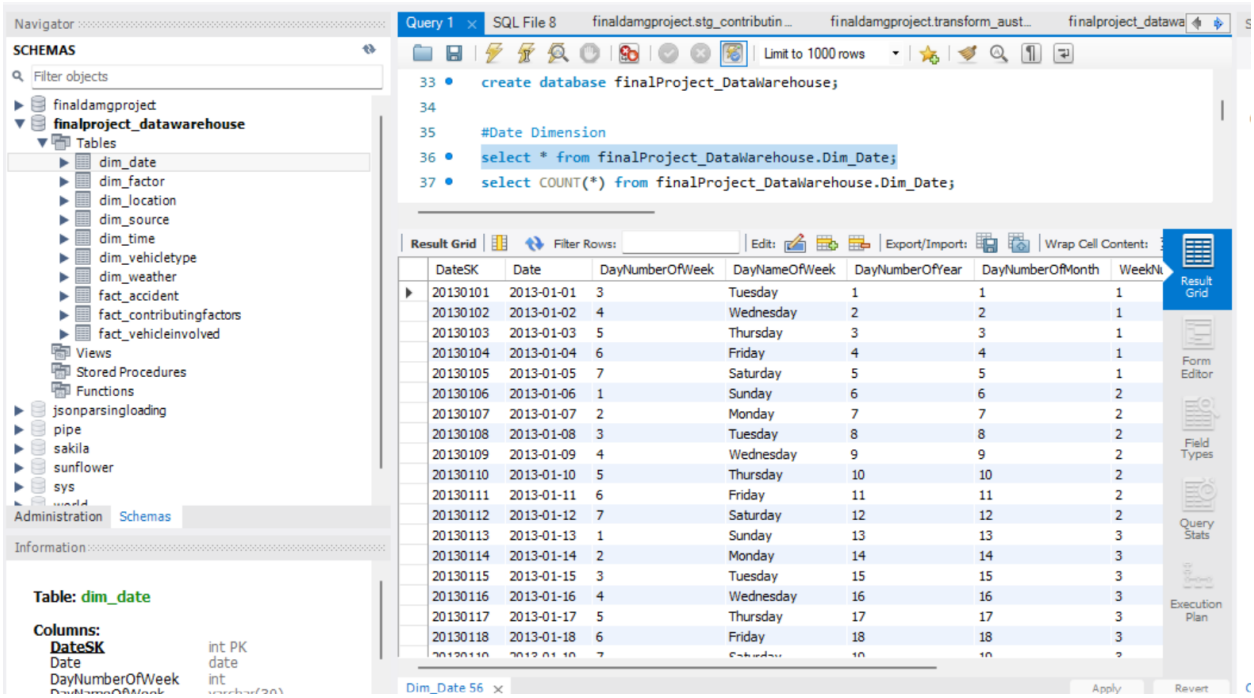
- FACT_Accident: Central fact table with AccidentSK, NumberofVehiclesInvolved, MotoristInjured, MotoristKilled, PedestriansInjured, PedestriansKilled, CyclistInjured, CyclistKilled, TotalInjured, TotalFatalities, DateSK, TimeSK, LocationSK, WeatherSK, SourceSK, DI_CreatedDate, DI_WorkFlowFileName.
- FACT_VehicleInvolved: Contains VehicleInvolvedSK, VehicleTypeSK, AccidentSK, DI_CreatedDate, DI_WorkFlowFileName.
- FACT_ContributingFactors: Includes ContributingFactorSK, FactorSK, AccidentSK, DI_CreatedDate, DI_WorkFlowFileName.

Finalproject_datawarehouse:

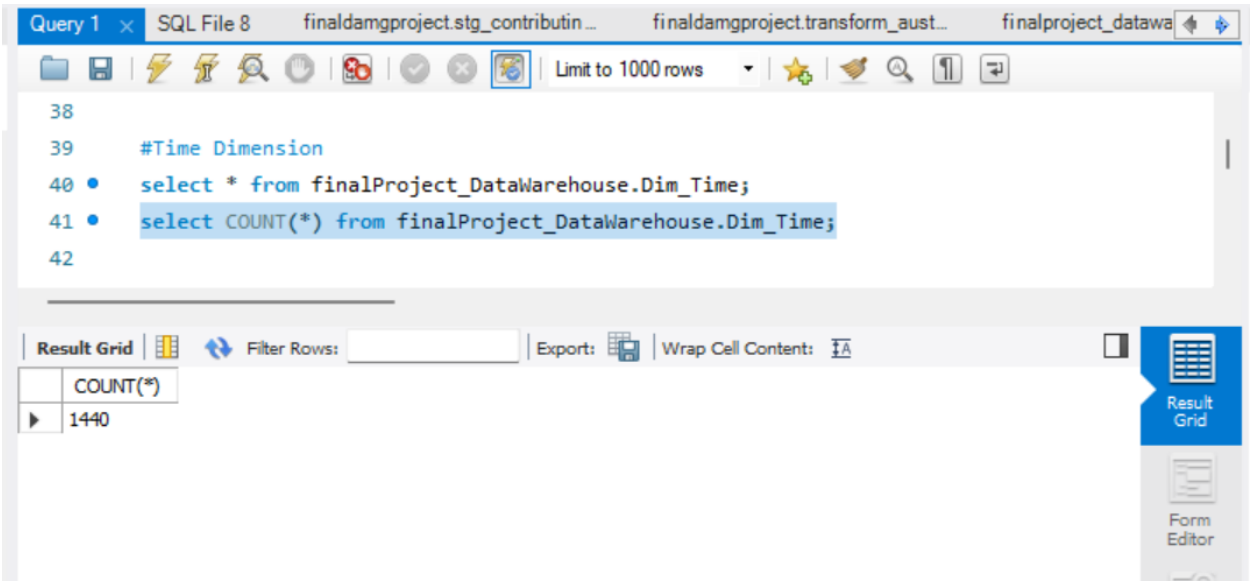
- Time Dimension



select * from finalProject_DataWarehouse.Dim_Time;



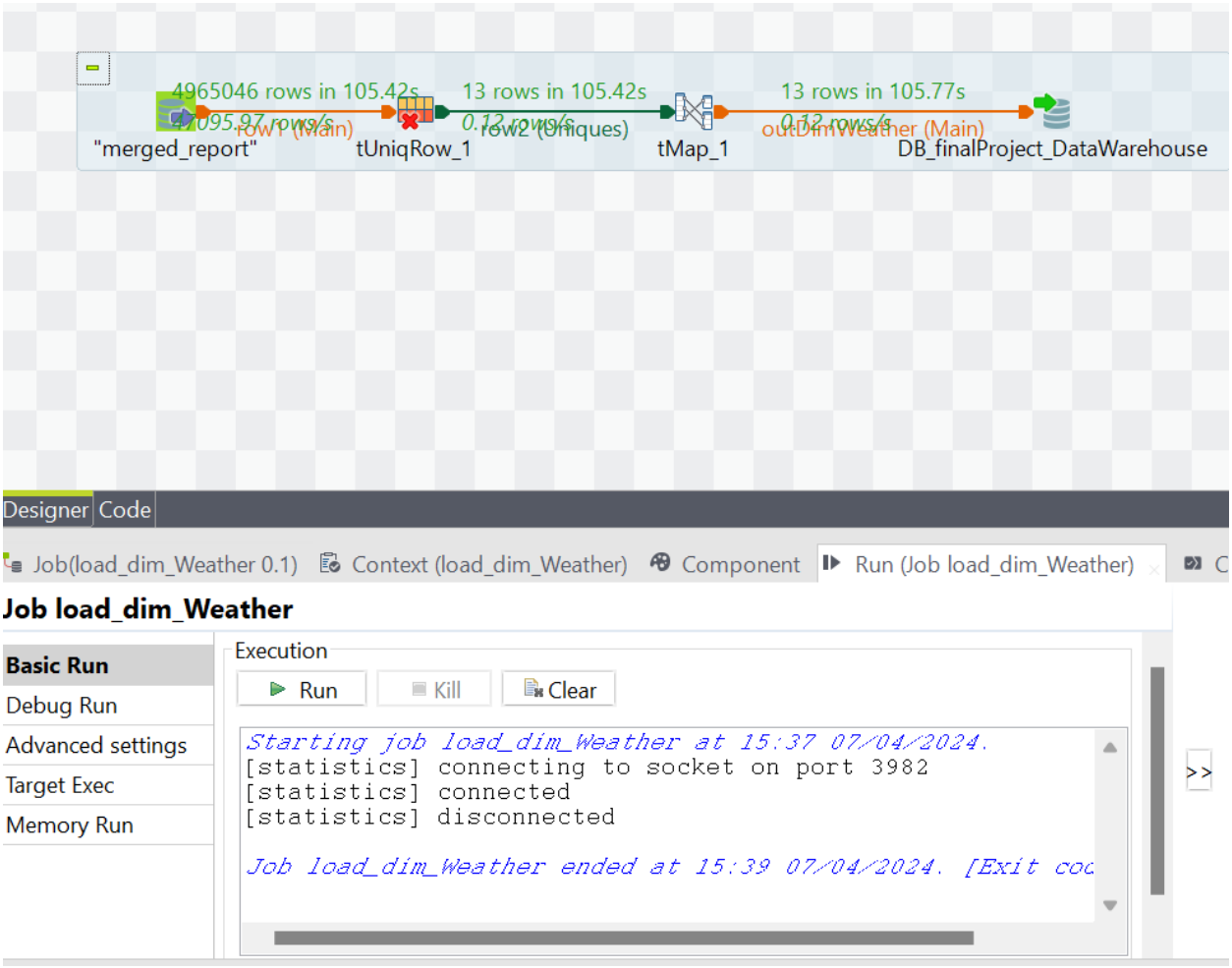
select COUNT(*) from finalProject_DataWarehouse.Dim_Time;



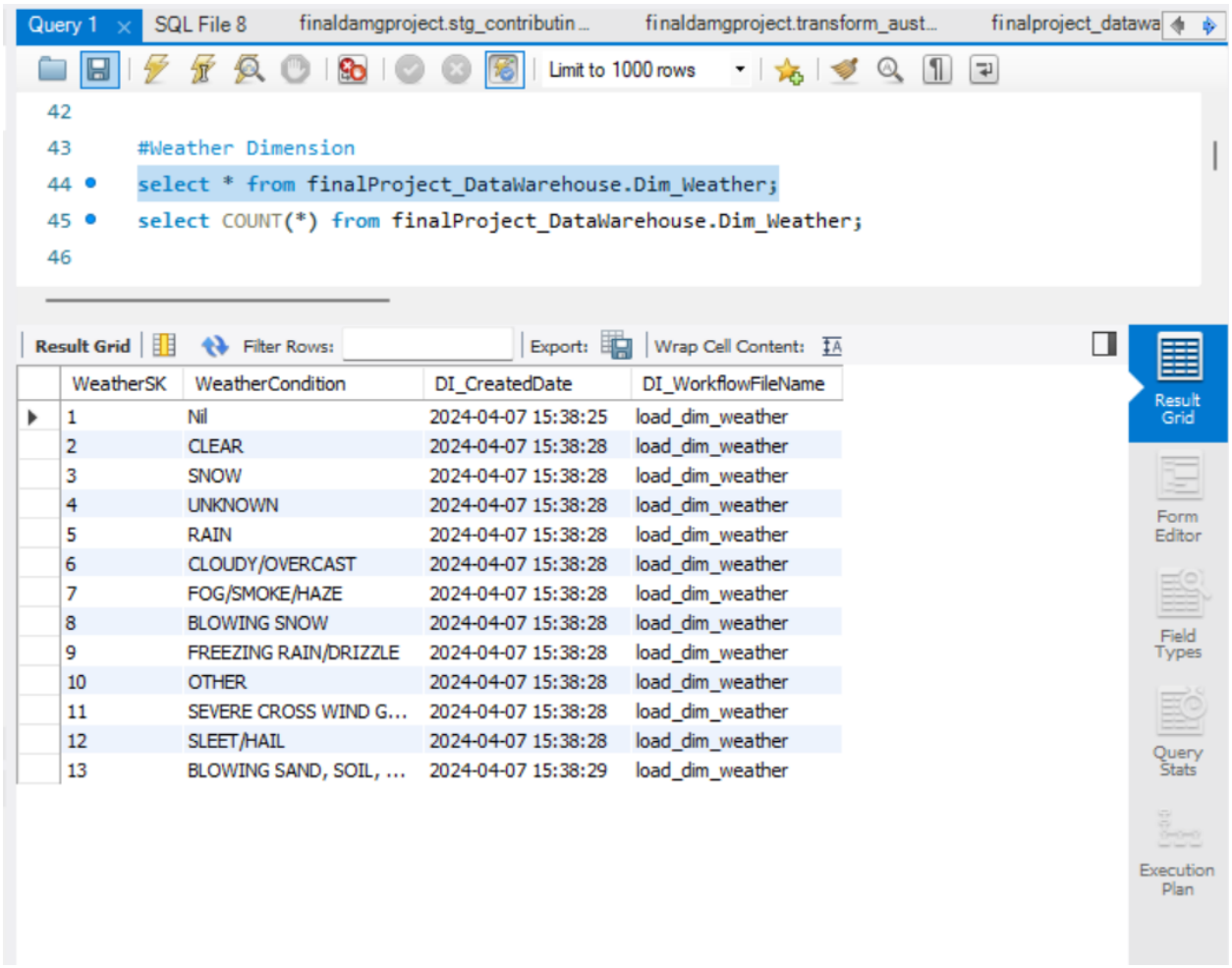

```
CREATE TABLE `dim_time` (
  `TimeSK` varchar(15) NOT NULL,
  `TimeOfTheDay` varchar(15) DEFAULT NULL,
  `Hour` int DEFAULT NULL,
  `DI_CreatedDate` datetime DEFAULT NULL,
  `DI_WorkflowFileName` varchar(50) DEFAULT NULL,
```

```
  PRIMARY KEY (`TimeSK`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci
```

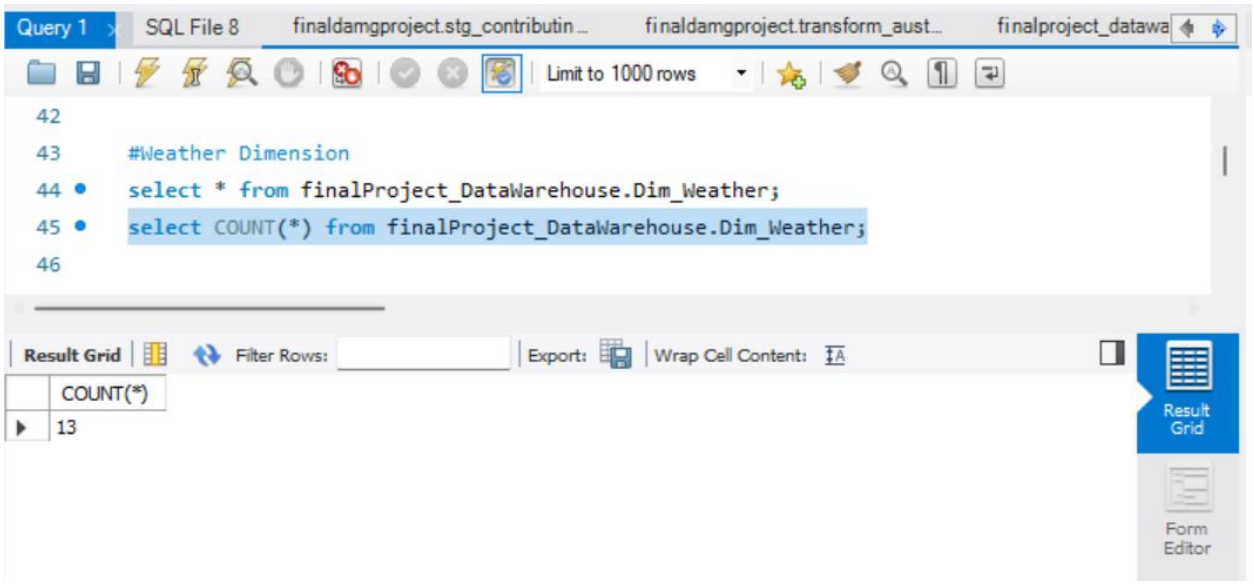
- Weather Dimension



select * from finalProject_DataWarehouse.Dim_Weather;

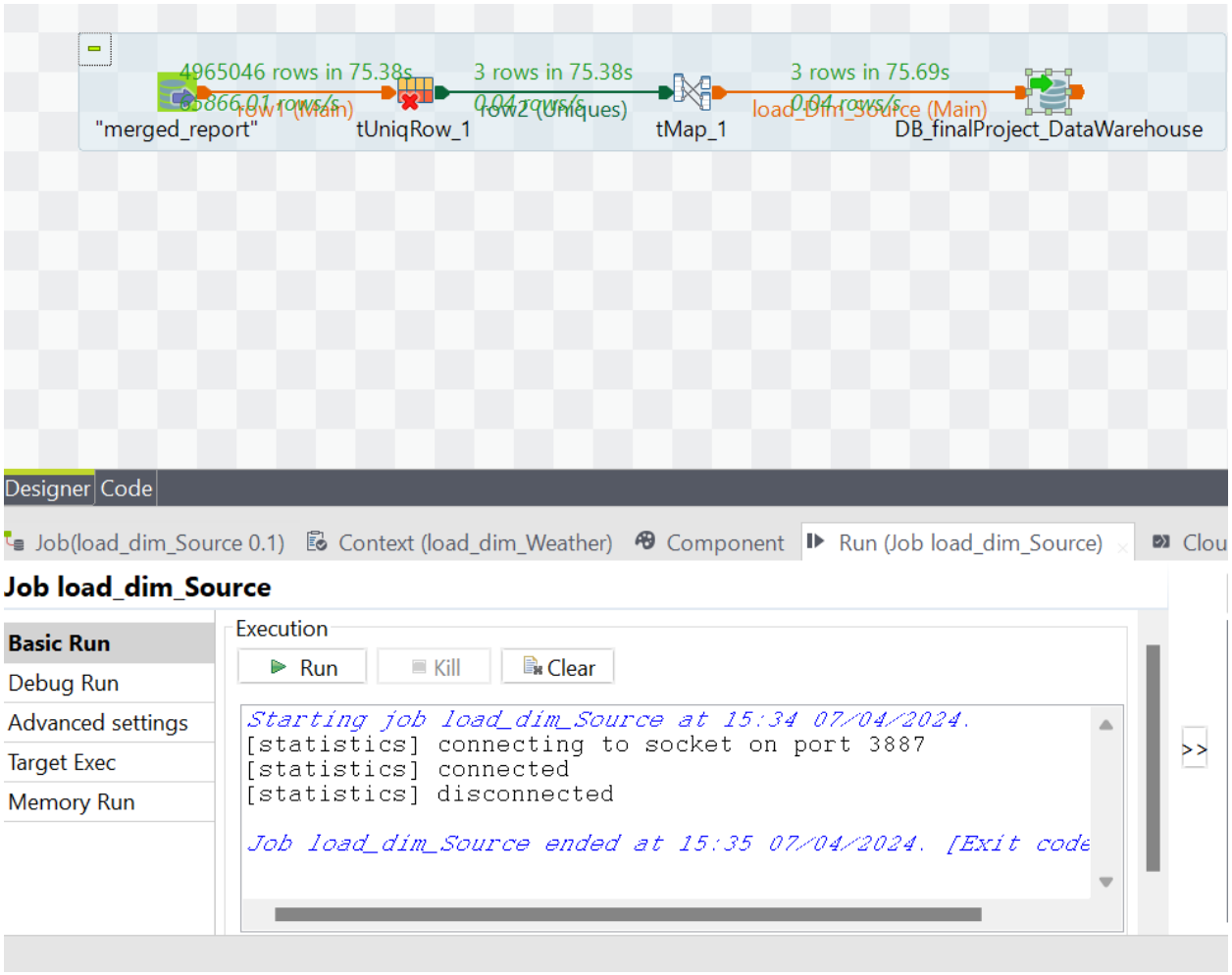


select COUNT(*) from finalProject_DataWarehouse.Dim_Weather;

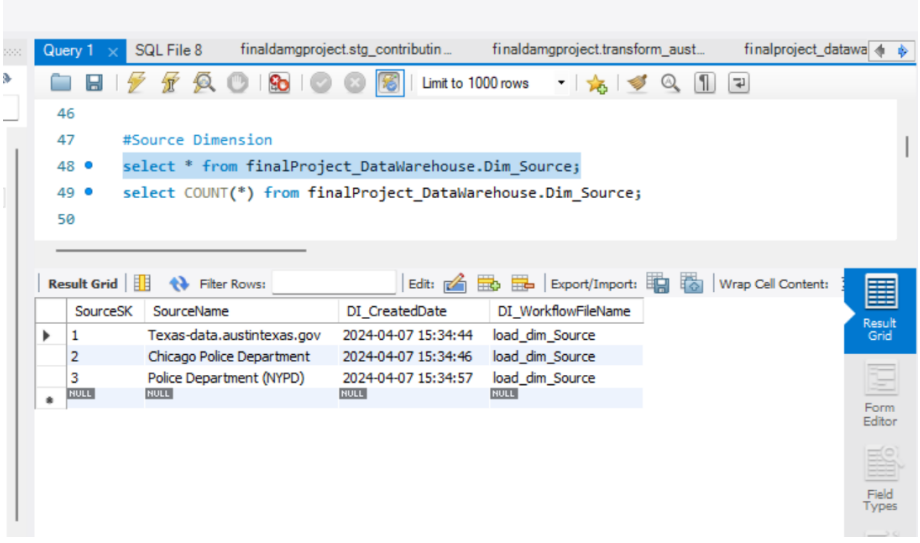



```
CREATE TABLE `dim_weather` (
  `WeatherSK` int DEFAULT NULL,
  `WeatherCondition` varchar(100) DEFAULT NULL,
  `DI_CreatedDate` datetime DEFAULT NULL,
  `DI_WorkflowFileName` varchar(20) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci
```

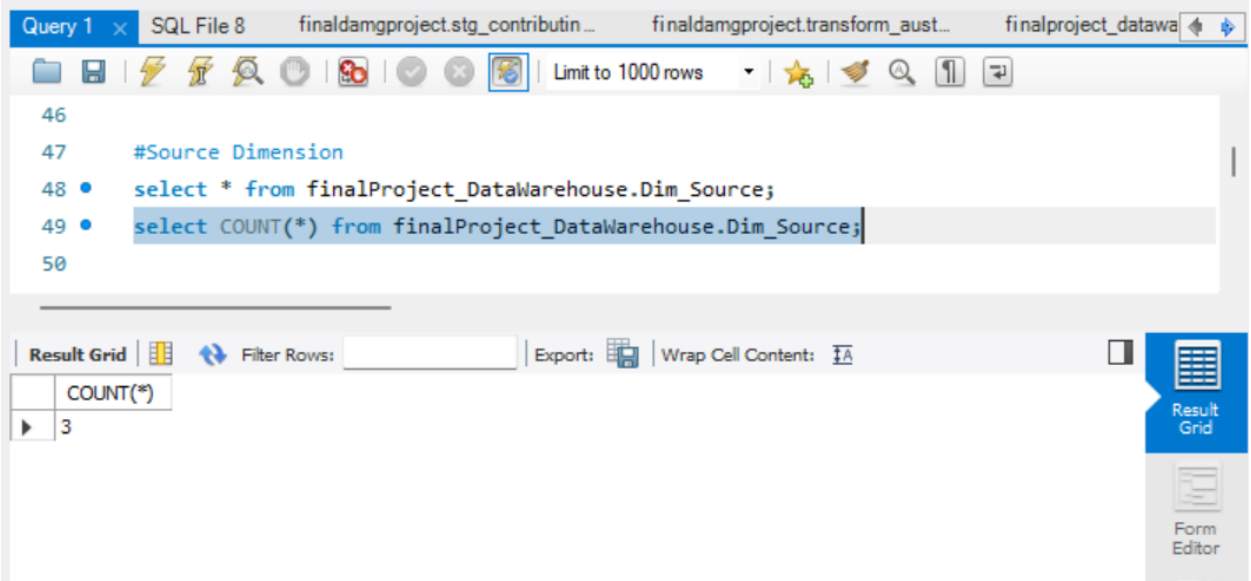
- Source Dimension



select * from finalProject_DataWarehouse.Dim_Source;

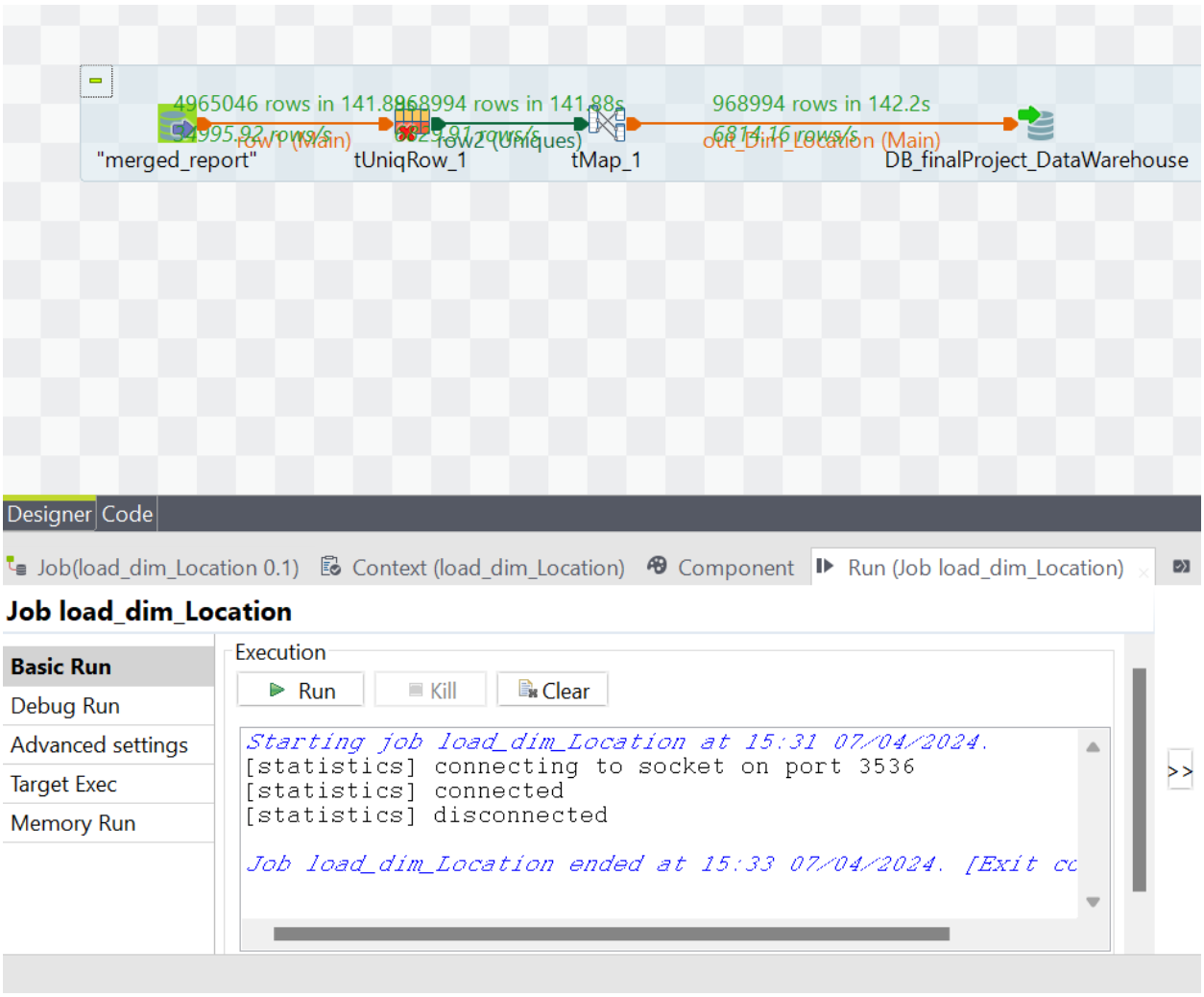


select COUNT(*) from finalProject_DataWarehouse.Dim_Source;

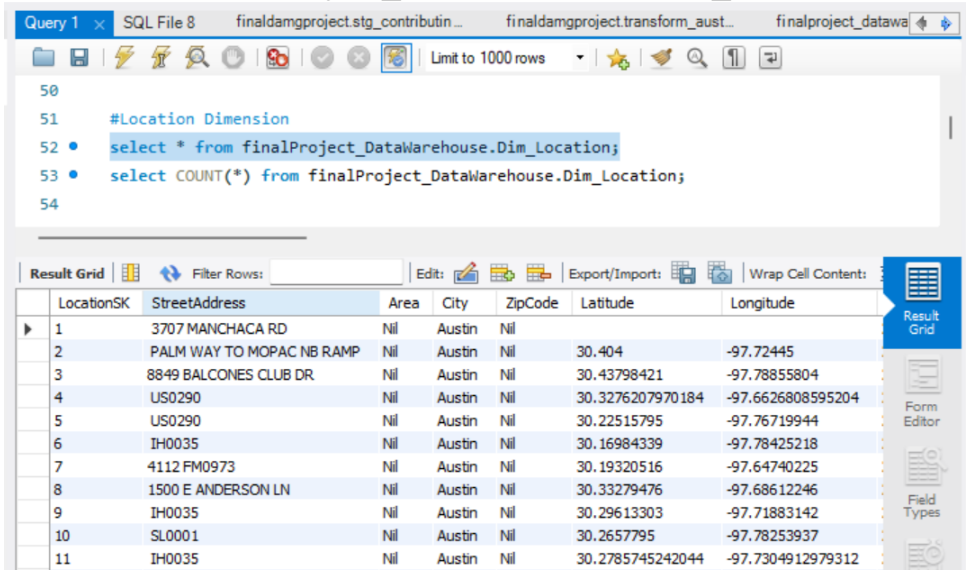



```
CREATE TABLE `dim_source` (
  `SourceSK` int NOT NULL,
  `SourceName` varchar(50) DEFAULT NULL,
  `DI_CreatedDate` datetime DEFAULT NULL,
  `DI_WorkflowFileName` varchar(50) DEFAULT NULL,
  PRIMARY KEY (`SourceSK`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
COLLATE=utf8mb4_0900_ai_ci
```

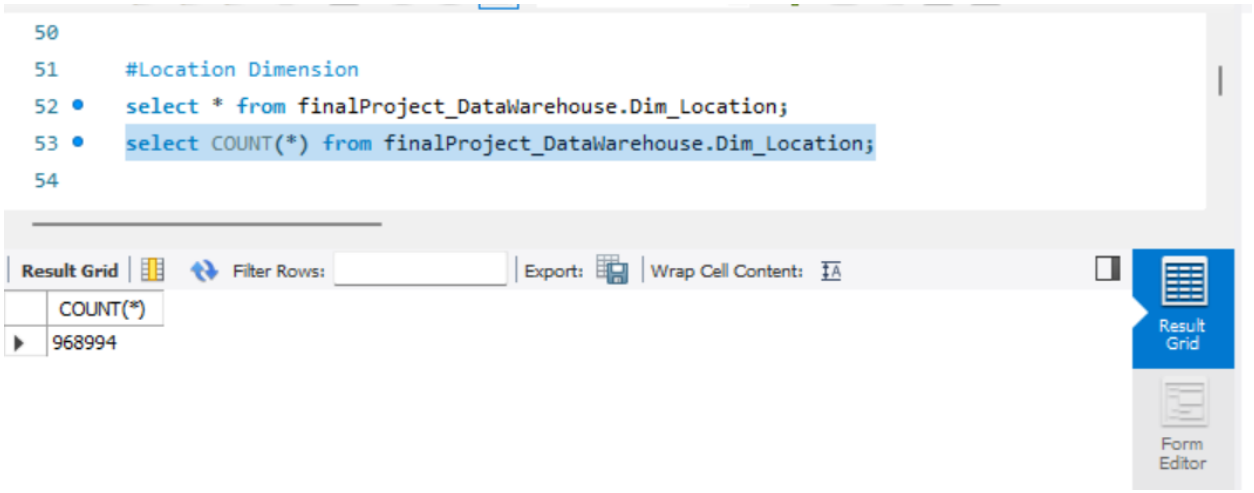
- Location Dimension



select * from finalProject_DataWarehouse.Dim_Location;

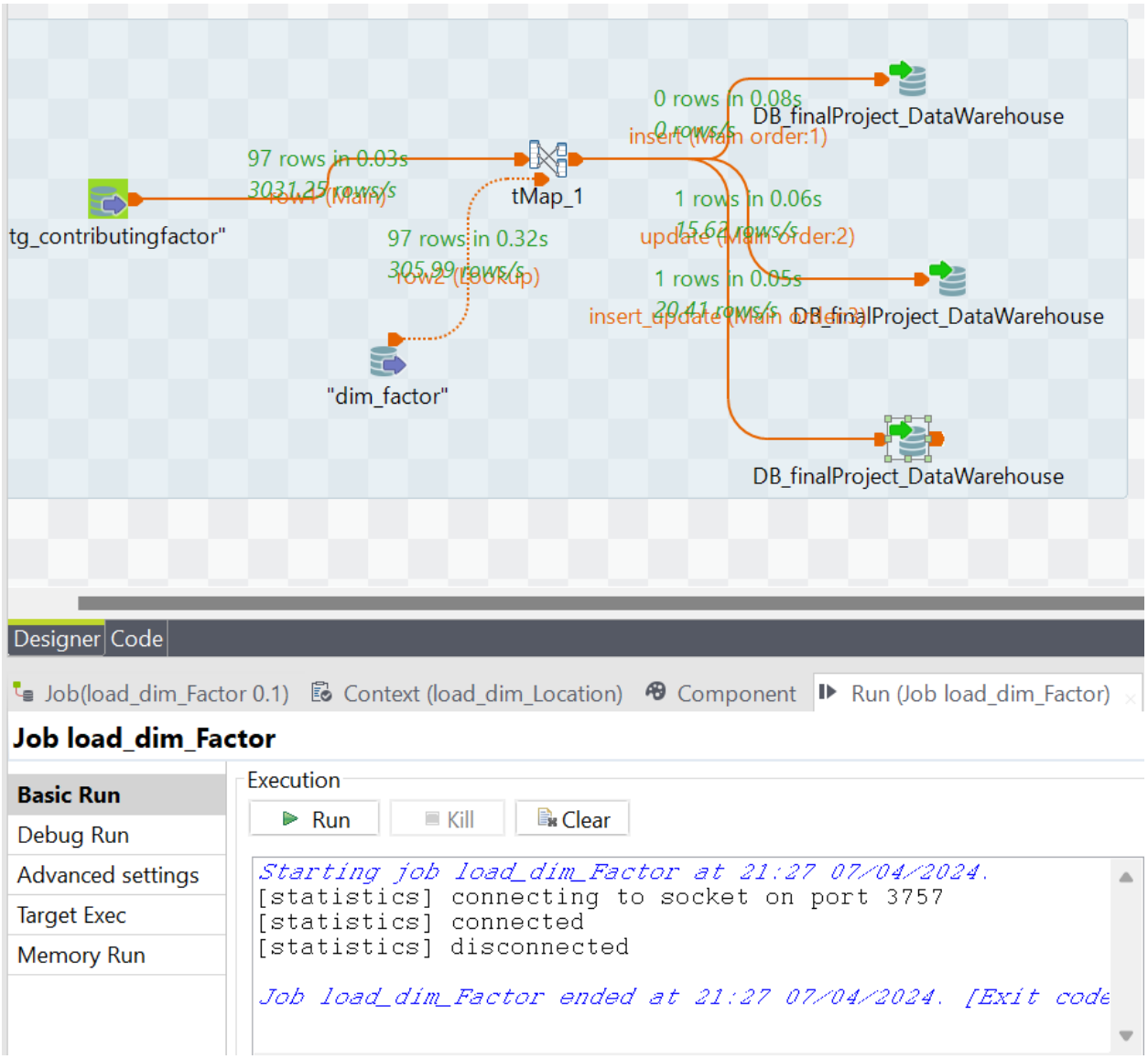


select COUNT(*) from finalProject_DataWarehouse.Dim_Location;




```
CREATE TABLE `dim_location` (
  `LocationSK` int NOT NULL,
  `StreetAddress` varchar(500) DEFAULT NULL,
  `Area` varchar(20) DEFAULT NULL,
  `City` varchar(20) DEFAULT NULL,
  `ZipCode` varchar(10) DEFAULT NULL,
  `Latitude` varchar(20) DEFAULT NULL,
  `Longitude` varchar(20) DEFAULT NULL,
  `DI_CreatedDate` datetime DEFAULT NULL,
  `DI_WorkflowFileName` varchar(50) DEFAULT NULL,
  PRIMARY KEY (`LocationSK`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
COLLATE=utf8mb4_0900_ai_ci
```

- Factor Dimension



select * from finalProject_DataWarehouse.Dim_Factor;

```
55      #Factor Dimension
56 •    select * from finalProject_DataWarehouse.Dim_Factor;
57 •    select COUNT(*) from finalProject_DataWarehouse.Dim_Factor;
58
59      #Factor VehicleType
60 •    select * from finalProject_DataWarehouse.Dim_VehicleType;
61 •    select COUNT(*) from finalProject_DataWarehouse.Dim_VehicleType;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

FactorSK	FactorCode	FactorDescription	StartDate	EndDate	ValidityIndicator
93	130	Pavement Defective	2024-04-07	NULL	Y
94	131	WEATHER	2024-04-07	NULL	Y
95	132	UNABLE TO DETERMINE	2024-04-07	NULL	Y
96	133	Driver Inexperience	2024-04-07	NULL	Y
97	134	Runaway Vehicle1	2024-04-07	2024-0...	N

Result Grid

select COUNT(*) from finalProject_DataWarehouse.Dim_Factor;

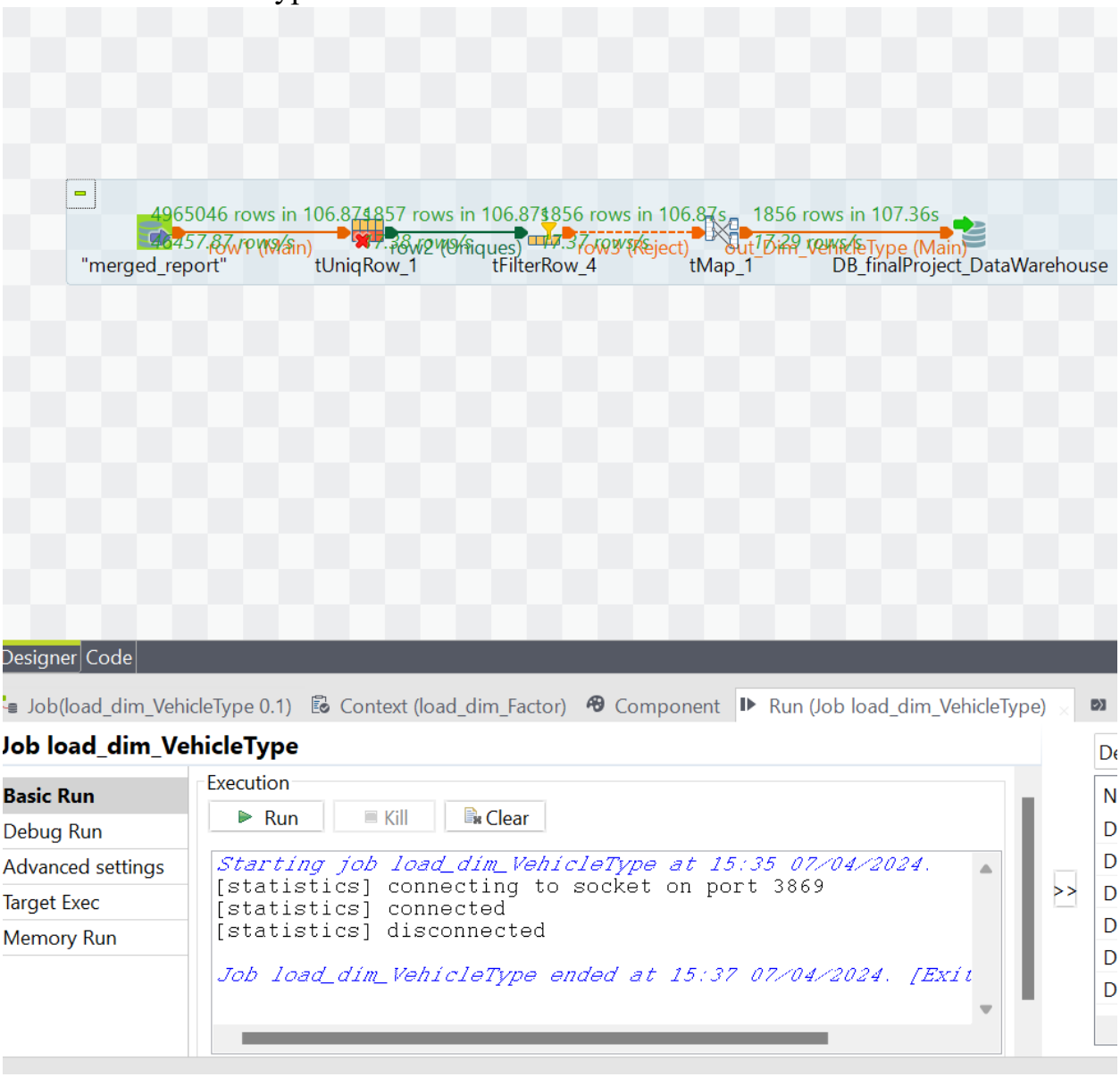
```
57 •    select COUNT(*) from finalProject_DataWarehouse.Dim_Factor;
58
59      #Factor VehicleType
60 •    select * from finalProject_DataWarehouse.Dim_VehicleType;
61 •    select COUNT(*) from finalProject_DataWarehouse.Dim_VehicleTyp
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

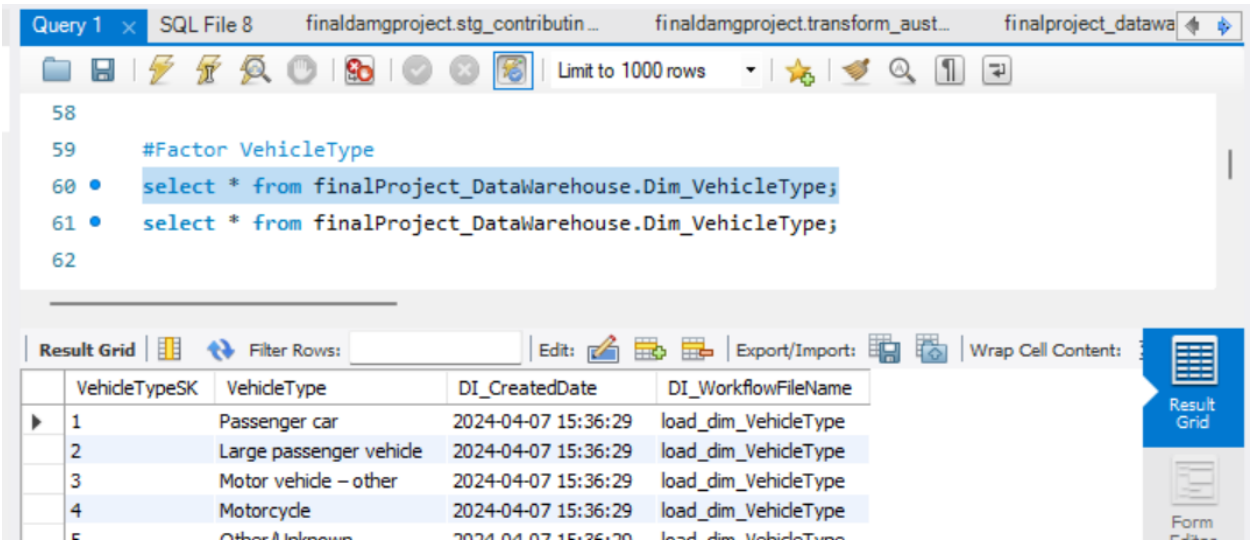
COUNT(*)
98

```
CREATE TABLE `dim_factor` (
  `FactorSK` int NOT NULL AUTO_INCREMENT,
  `FactorCode` int NOT NULL,
  `FactorDescription` varchar(1000) NOT NULL,
  `StartDate` date DEFAULT NULL,
  `EndDate` date DEFAULT NULL,
  `ValidityIndicator` varchar(2) DEFAULT NULL,
  PRIMARY KEY (`FactorSK`)
) ENGINE=InnoDB AUTO_INCREMENT=99 DEFAULT
CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci
```

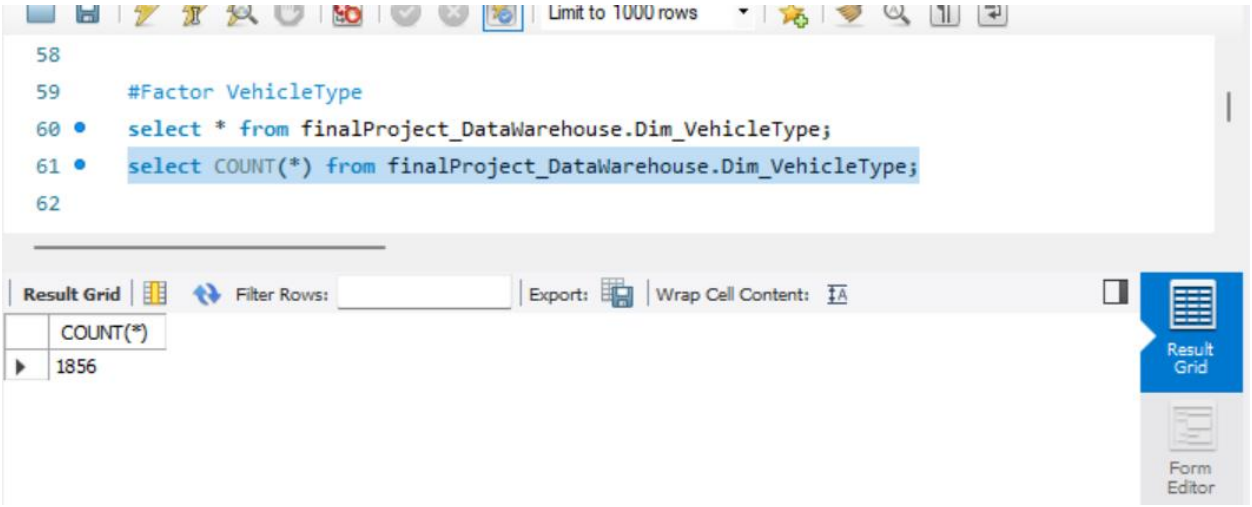
- Dimension VehicleType



select * from finalProject_DataWarehouse.Dim_VehicleType;

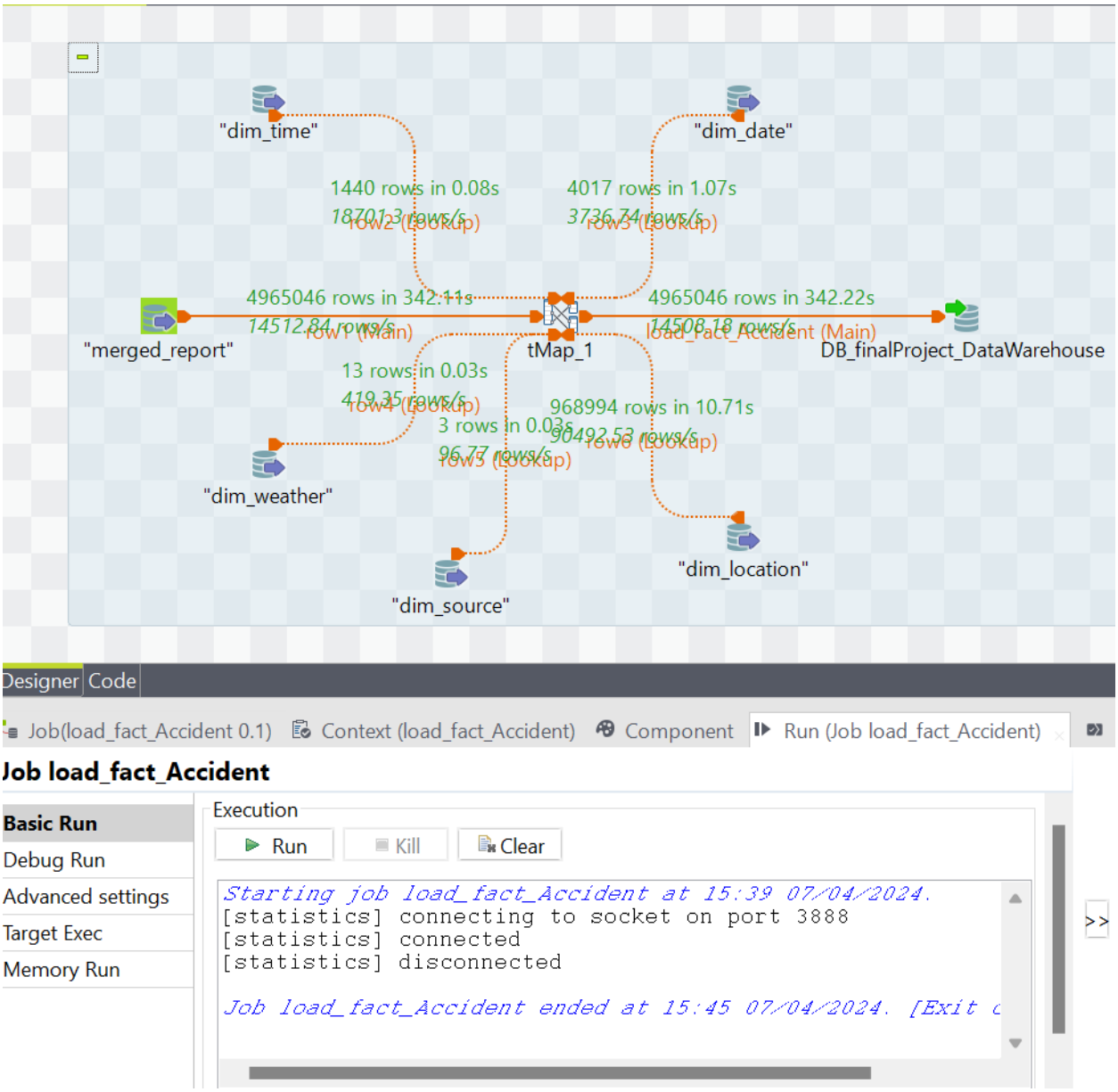


select * from finalProject_DataWarehouse.Dim_VehicleType;

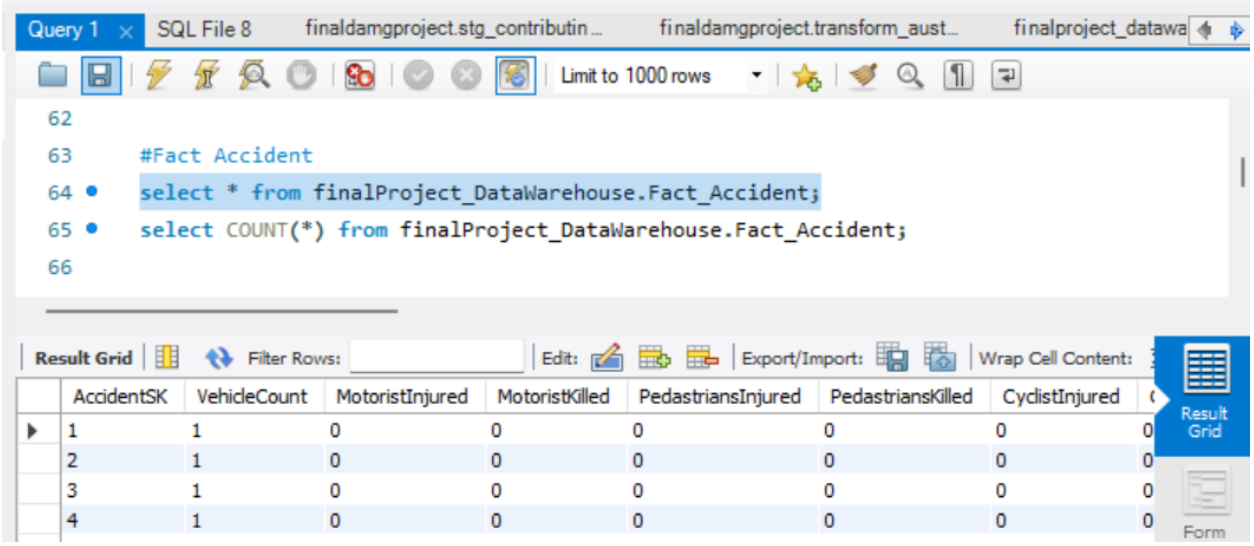



```
CREATE TABLE `dim_vehicletype` (
  `VehicleTypeSK` int NOT NULL,
  `VehicleType` varchar(250) DEFAULT NULL,
  `DI_CreatedDate` datetime DEFAULT NULL,
  `DI_WorkflowFileName` varchar(50) DEFAULT NULL,
  PRIMARY KEY (`VehicleTypeSK`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
COLLATE=utf8mb4_0900_ai_ci
```

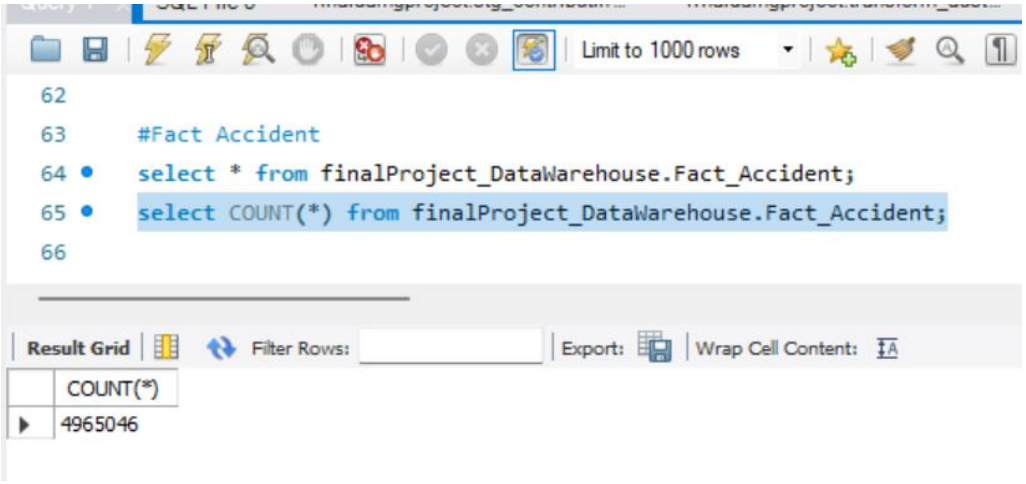
- Fact Accident



select * from finalProject_DataWarehouse.Fact_Accident;



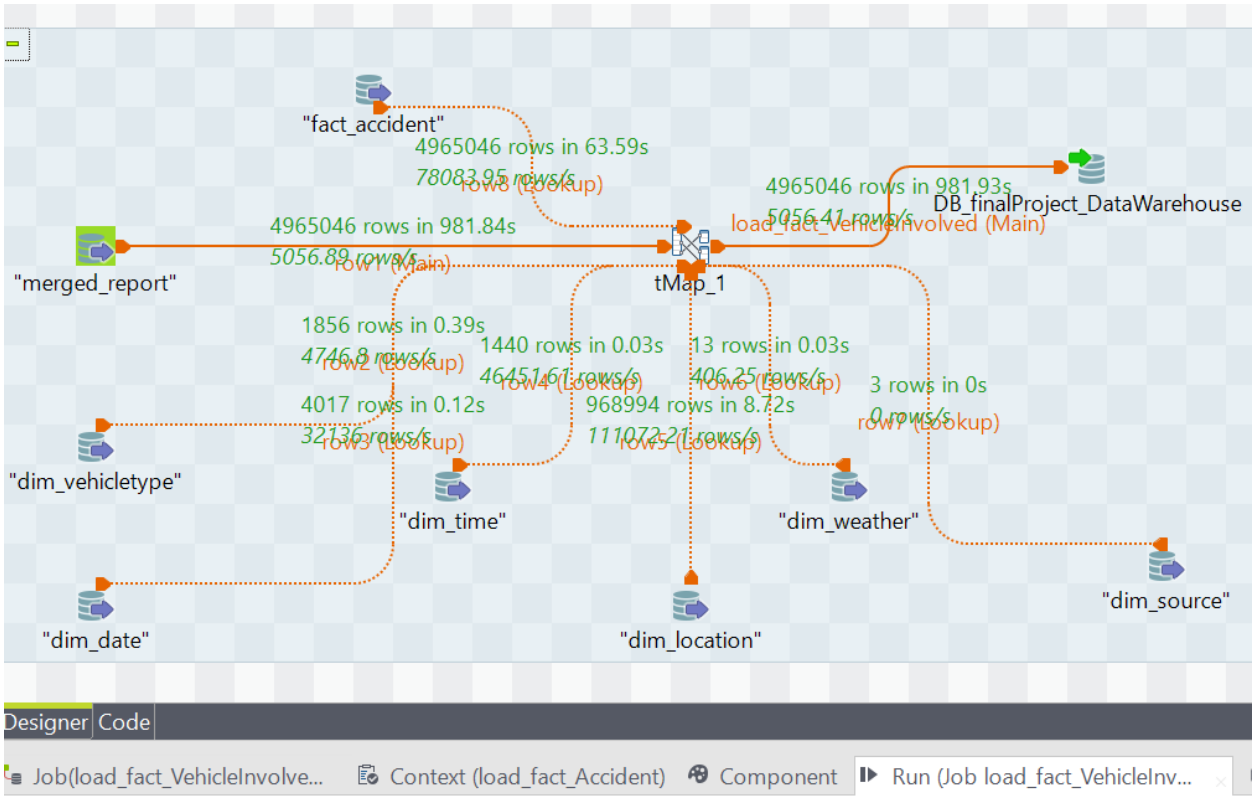
select COUNT(*) from finalProject_DataWarehouse.Fact_Accident;

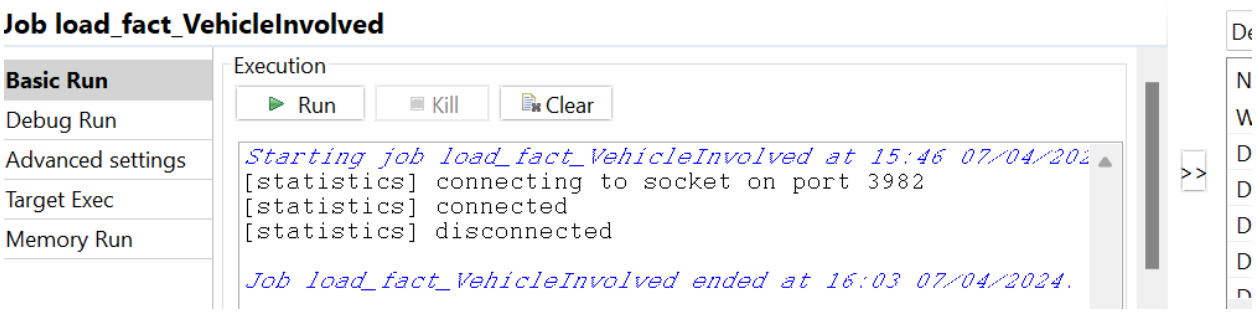



```
CREATE TABLE `fact_accident` (
  `AccidentSK` int NOT NULL,
  `VehicleCount` int DEFAULT NULL,
  `MotoristInjured` int DEFAULT NULL,
  `MotoristKilled` int DEFAULT NULL,
  `PedastriansInjured` int DEFAULT NULL,
  `PedastriansKilled` int DEFAULT NULL,
  `CyclistInjured` int DEFAULT NULL,
  `CyclistKilled` int DEFAULT NULL,
  `TotalInjured` int DEFAULT NULL,
```

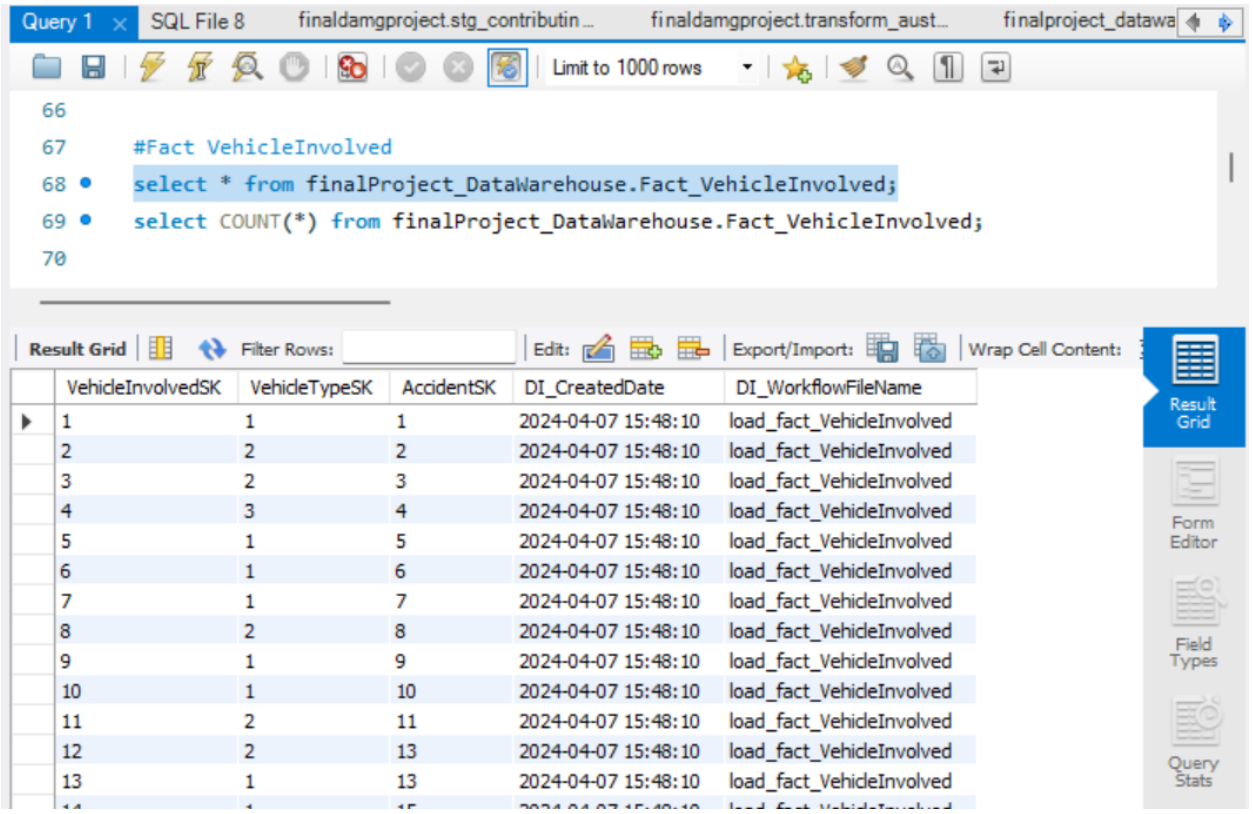
```
  `TotalKilled` int DEFAULT NULL,
  `DateSK` int NOT NULL,
  `TimeSK` varchar(15) NOT NULL,
  `LocationSK` int NOT NULL,
  `WeatherSK` int NOT NULL,
  `SourceSK` int NOT NULL,
  `DI_CreatedDate` datetime DEFAULT NULL,
  `DI_WorkflowFileName` varchar(50) DEFAULT NULL,
  PRIMARY KEY
(`AccidentSK`,`DateSK`,`TimeSK`,`LocationSK`,`WeatherSK`,`SourceSK`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
COLLATE=utf8mb4_0900_ai_ci
```

- Fact VehicleInvolved

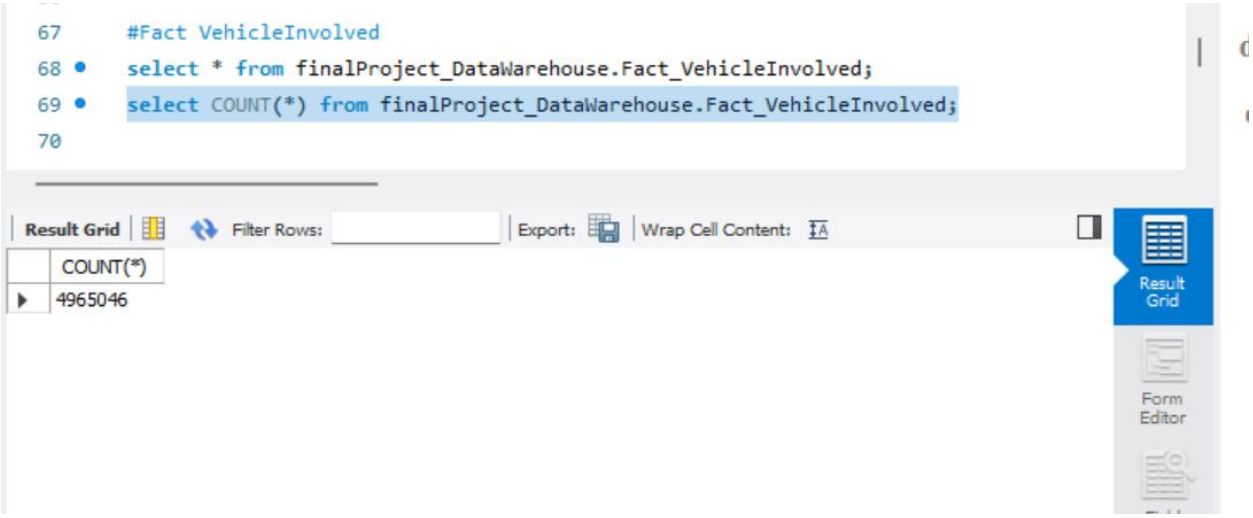




select * from finalProject_DataWarehouse.Fact_VehicleInvolved;



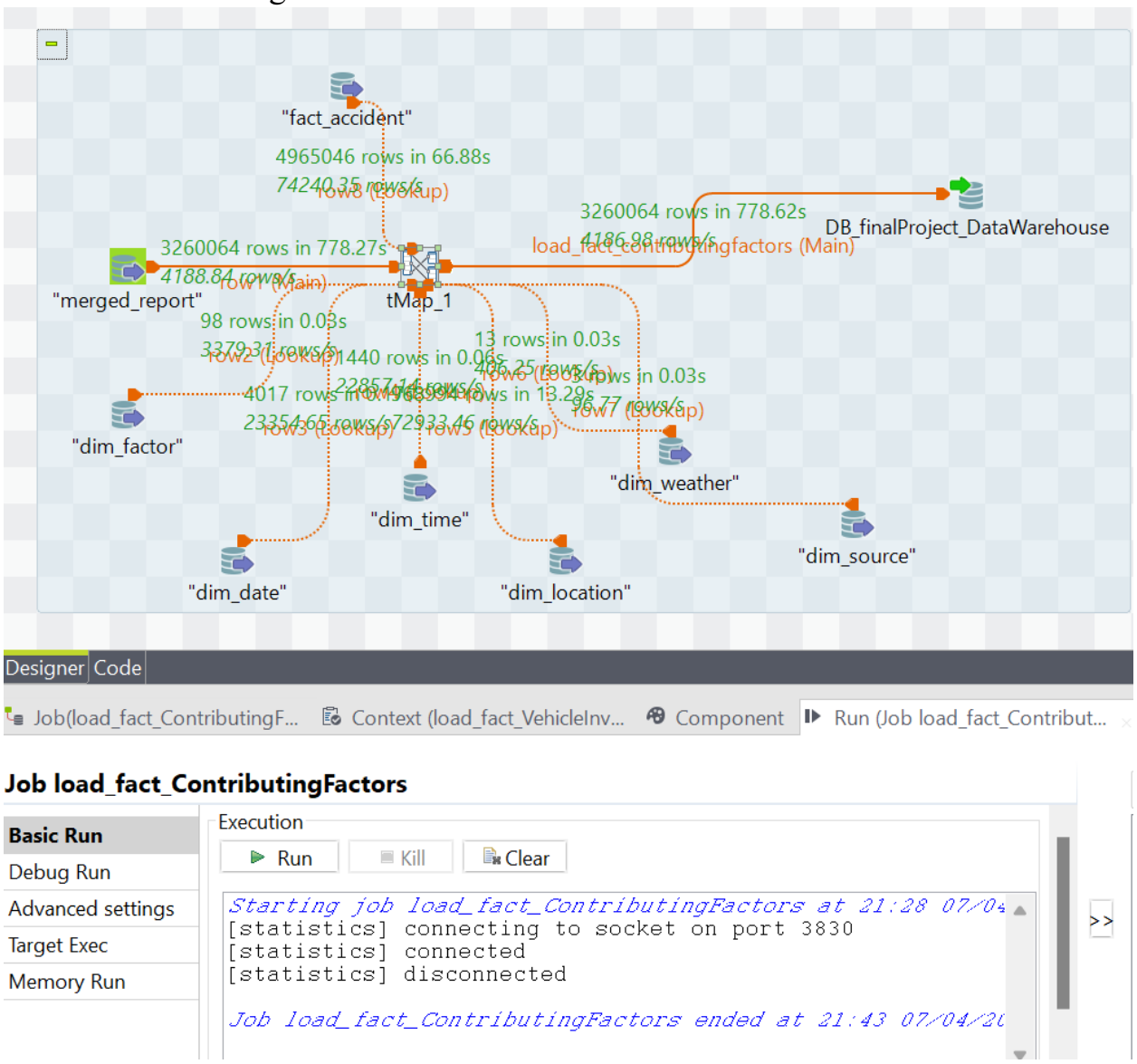
select COUNT(*) from finalProject_DataWarehouse.Fact_VehicleInvolved;



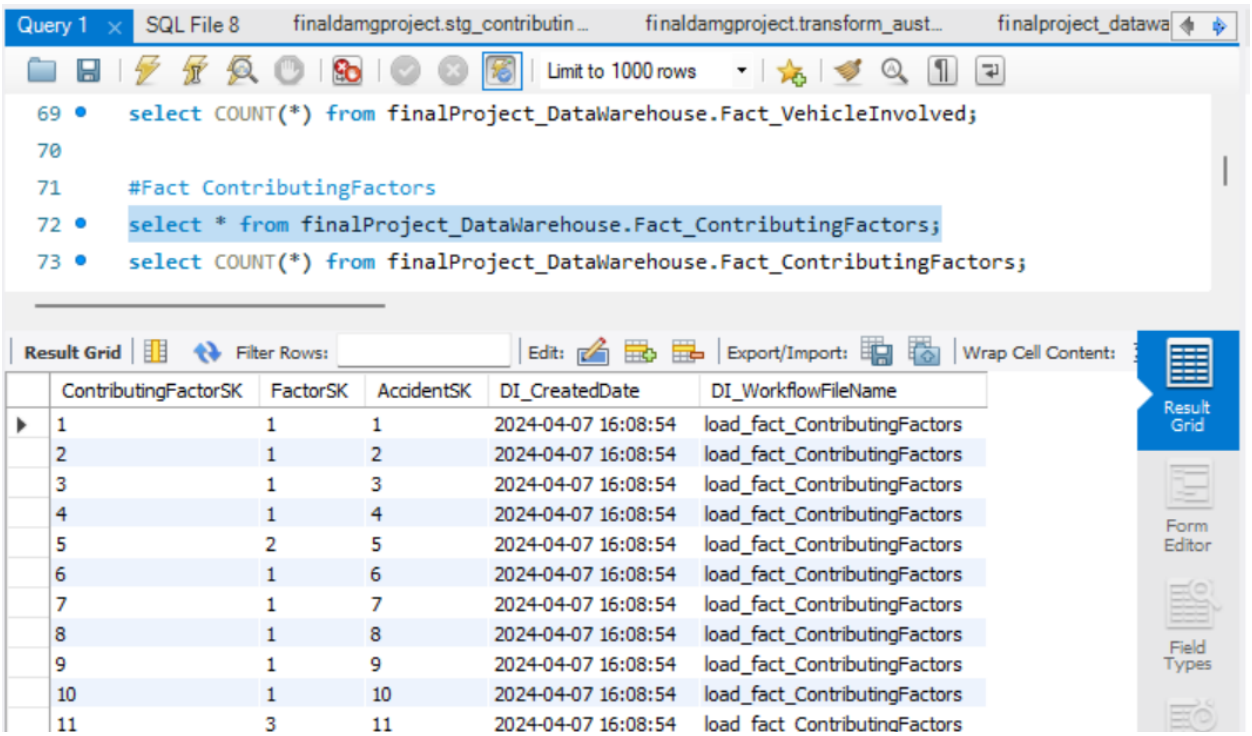

```
CREATE TABLE `fact_vehicleinvolved` (
  `VehicleInvolvedSK` int NOT NULL,
  `VehicleTypeSK` int NOT NULL,
  `AccidentSK` int NOT NULL,
  `DI_CreatedDate` datetime DEFAULT NULL,
  `DI_WorkflowFileName` varchar(50) DEFAULT NULL,
```

PRIMARY KEY (`VehicleInvolvedSK`,`VehicleTypeSK`,`AccidentSK`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
COLLATE=utf8mb4_0900_ai_ci

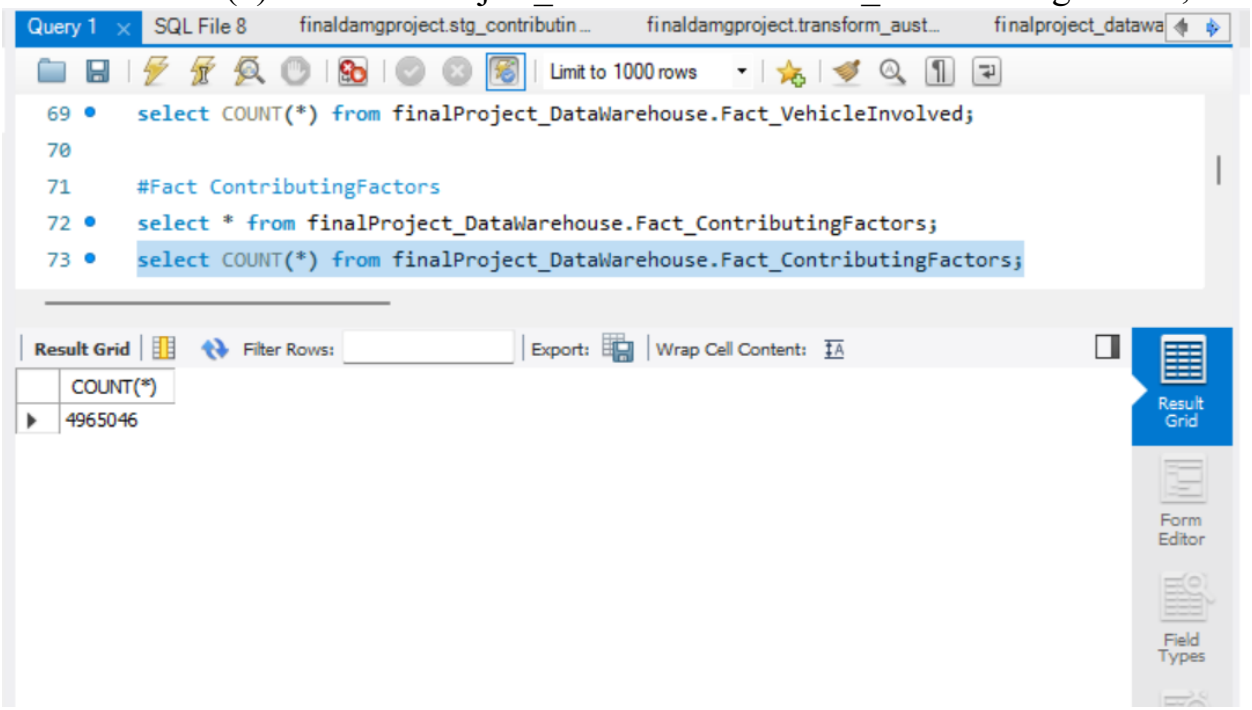
- Fact ContributingFactors



select * from finalProject_DataWarehouse.Fact_ContributingFactors;



select COUNT(*) from finalProject_DataWarehouse.Fact_ContributingFactors;



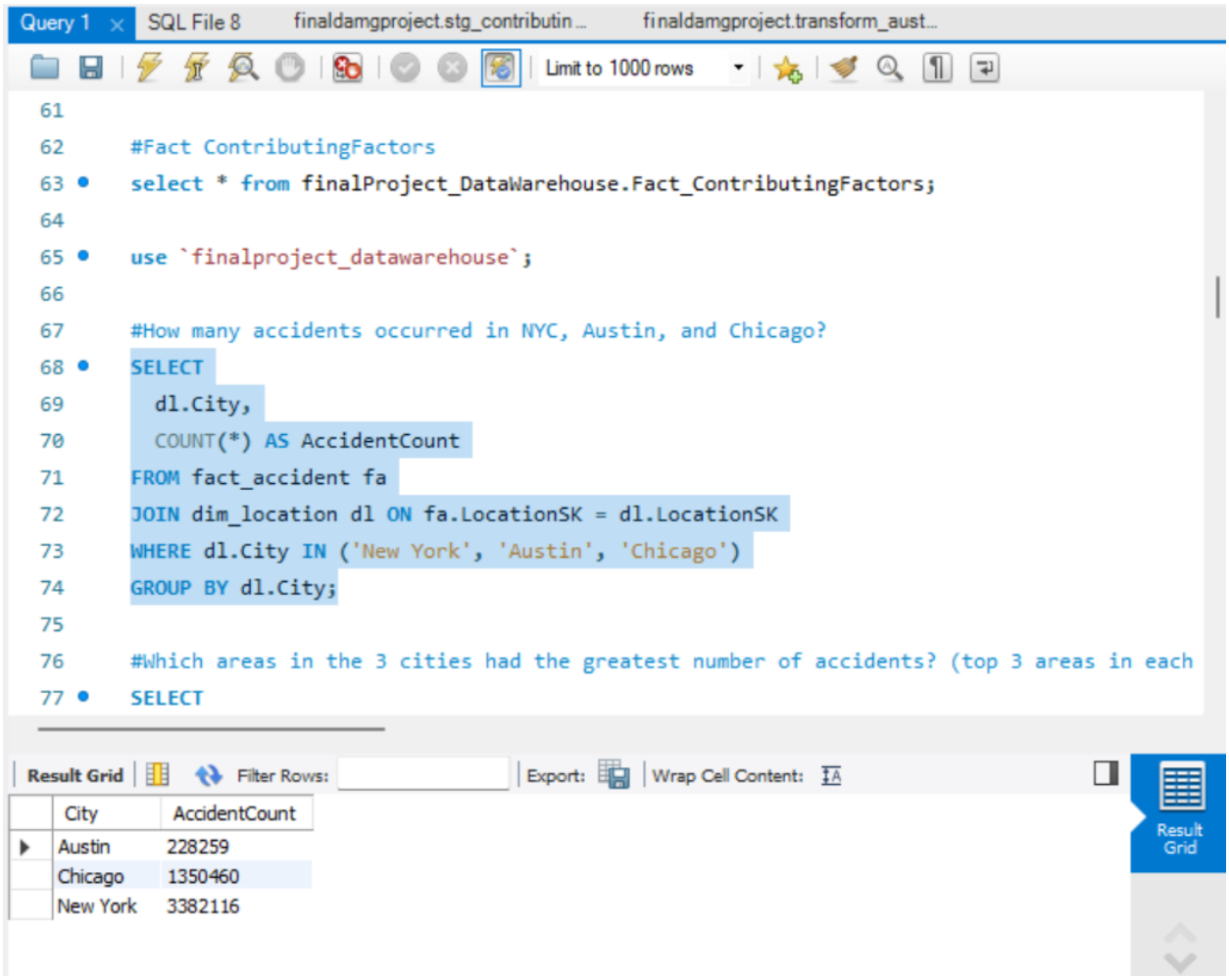

```
CREATE TABLE `fact_contributingfactors` (
  `ContributingFactorSK` int NOT NULL,
  `FactorSK` int NOT NULL,
  `AccidentSK` int NOT NULL,
  `DI_CreatedDate` datetime DEFAULT NULL,
  `DI_WorkflowFileName` varchar(50) DEFAULT NULL,
```

```
  PRIMARY KEY (`ContributingFactorSK`,`FactorSK`,`AccidentSK`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci
```

SQL Scripts to Validate Business Questions

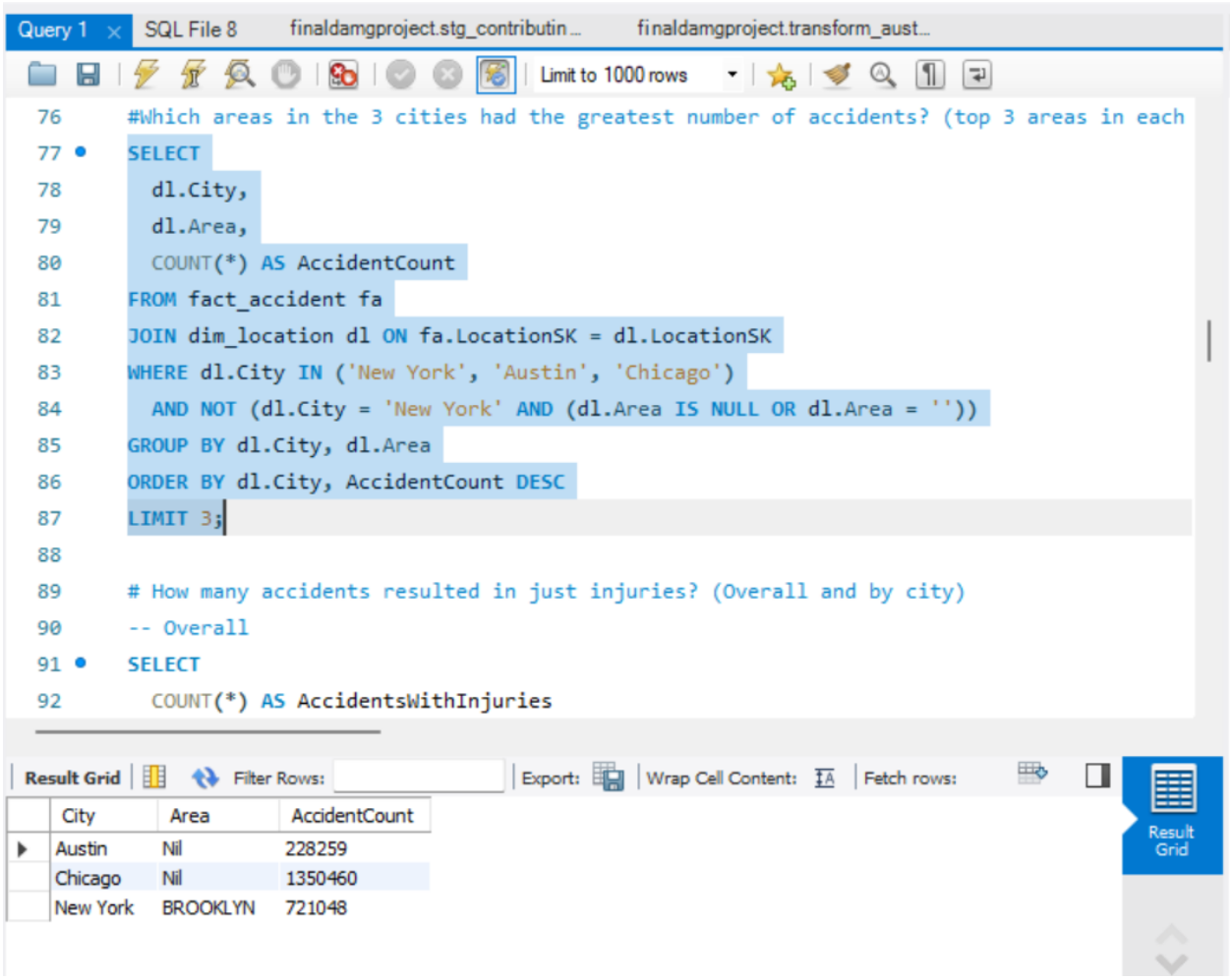
1. How many accidents occurred in NYC, Austin, and Chicago?

```
SELECT
  dl.City,
  COUNT(*) AS AccidentCount
FROM fact_accident fa
JOIN dim_location dl ON fa.LocationSK = dl.LocationSK
WHERE dl.City IN ('New York', 'Austin', 'Chicago')
GROUP BY dl.City;
```



2. Which areas in the 3 cities had the greatest number of accidents? (Top 3 areas in each city)

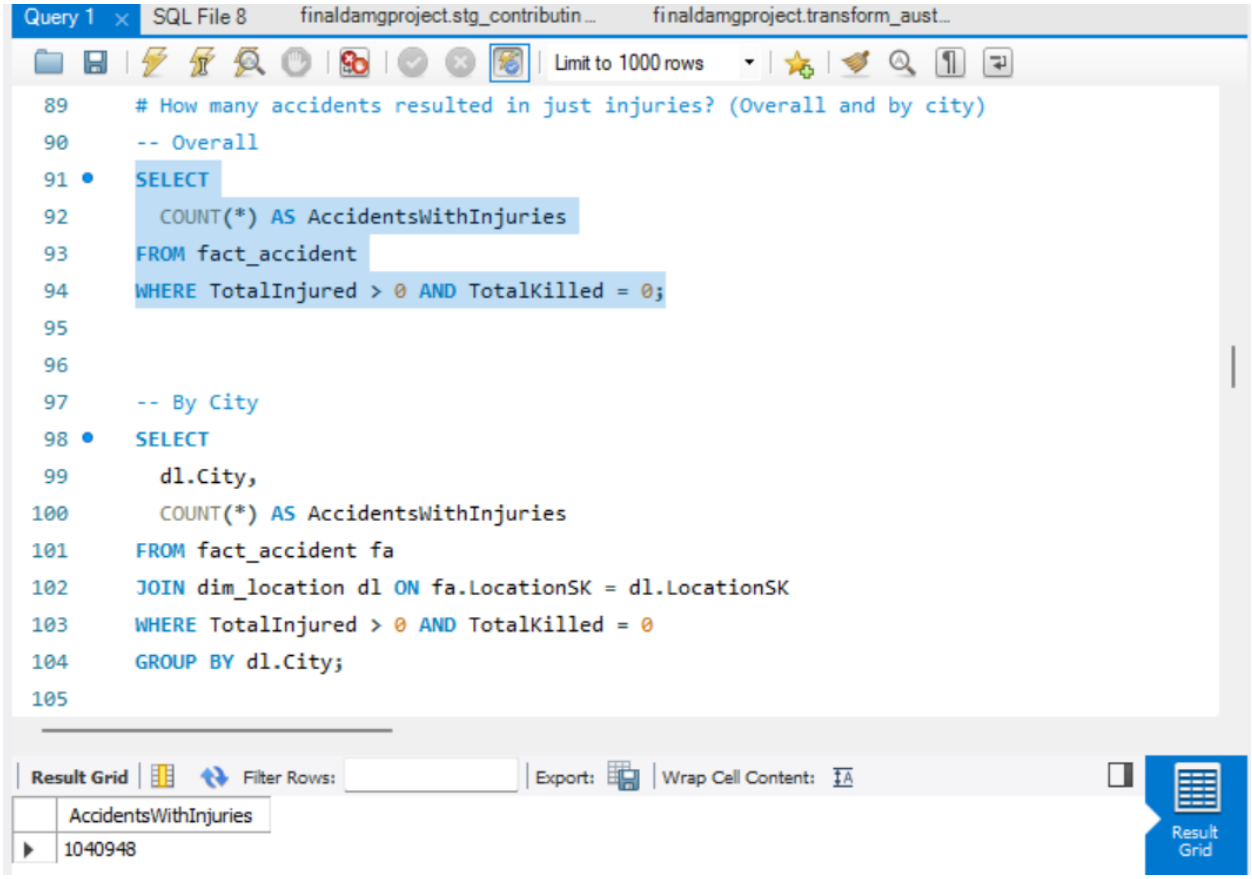
```
SELECT
  dl.City,
  dl.Area,
  COUNT(*) AS AccidentCount
FROM fact_accident fa
JOIN dim_location dl ON fa.LocationSK = dl.LocationSK
WHERE dl.City IN ('New York', 'Austin', 'Chicago')
  AND NOT (dl.City = 'New York' AND (dl.Area IS NULL OR dl.Area = ''))
GROUP BY dl.City, dl.Area
ORDER BY dl.City, AccidentCount DESC
LIMIT 3;
```

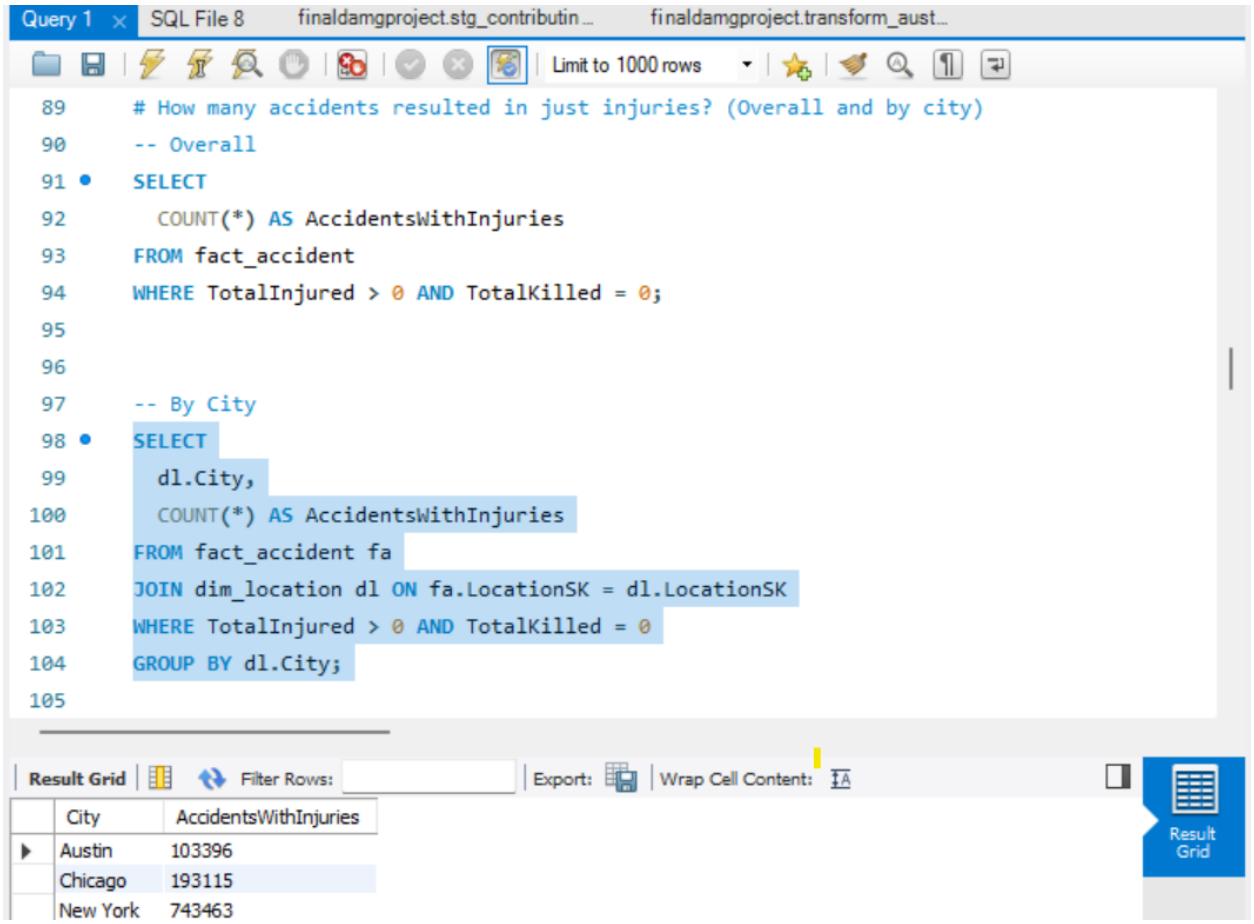


3. How many accidents resulted in just injuries? (Overall and by city)

```
-- Overall
SELECT
  COUNT(*) AS AccidentsWithInjuries
FROM fact_accident
WHERE TotalInjured > 0 AND TotalKilled = 0;


-- By City
SELECT
  dl.City,
  COUNT(*) AS AccidentsWithInjuries
FROM fact_accident fa
```

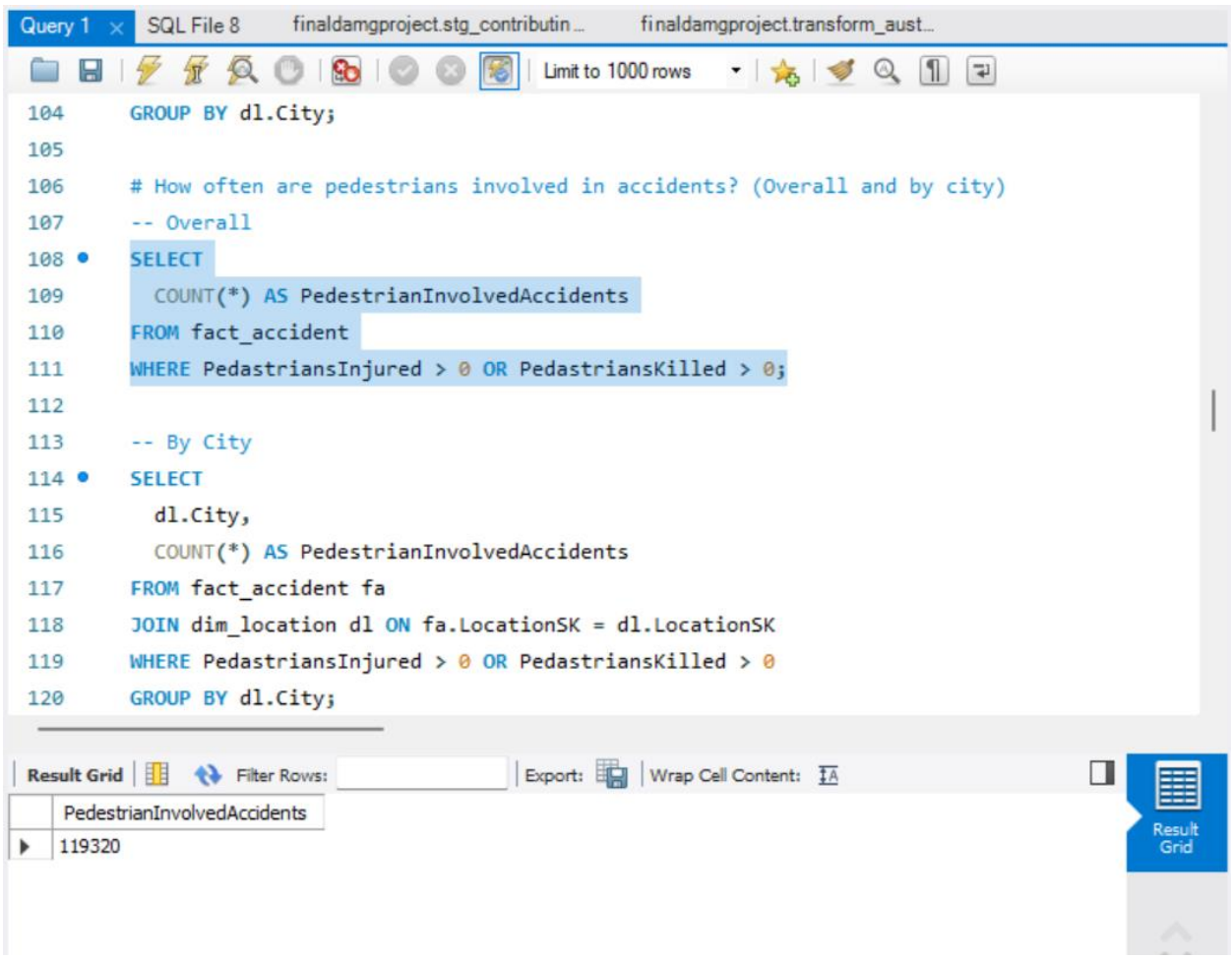
```
JOIN dim_location dl ON fa.LocationSK = dl.LocationSK
WHERE TotalInjured > 0 AND TotalKilled = 0
GROUP BY dl.City;
```

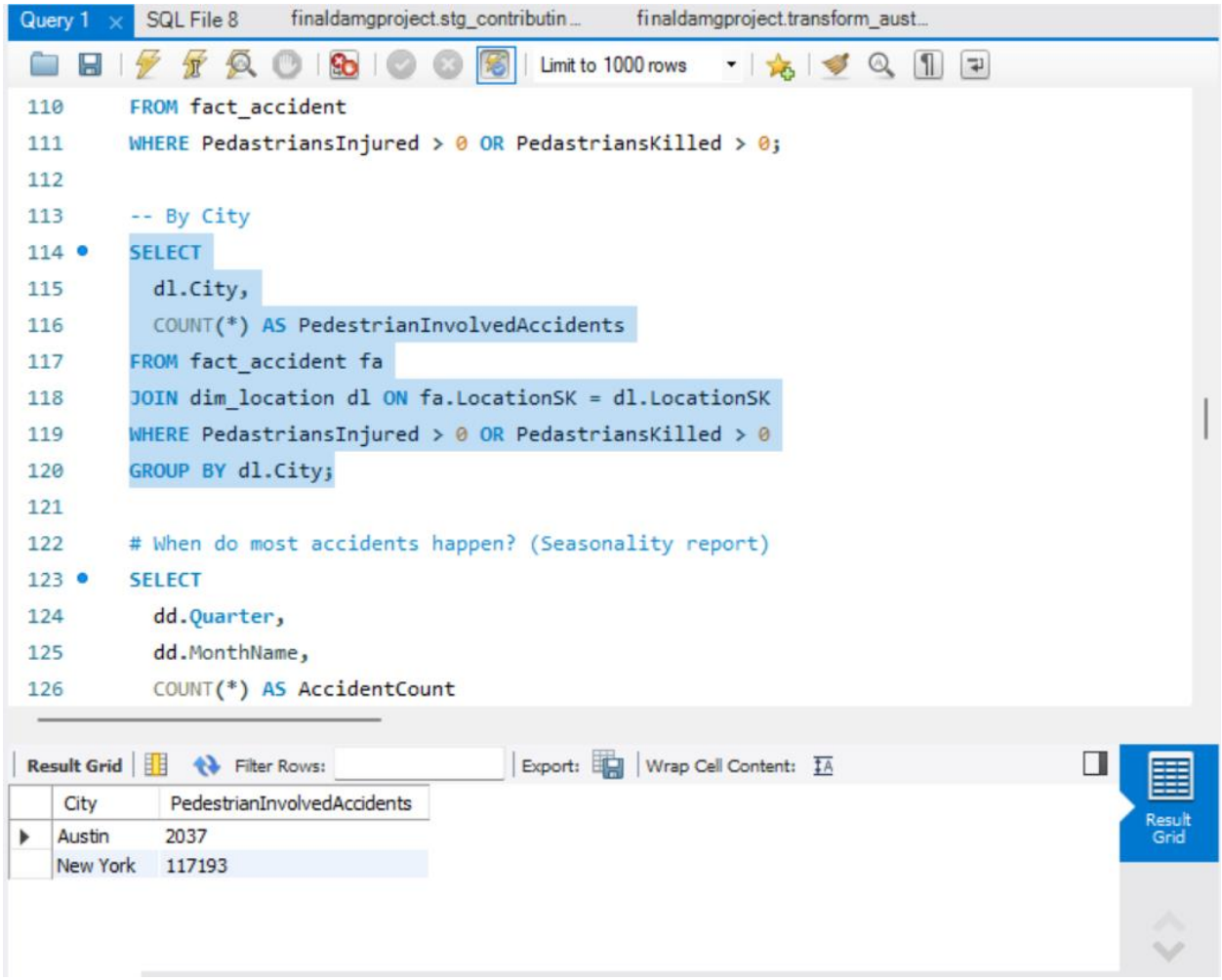




4. How often are pedestrians involved in accidents? (Overall and by city)

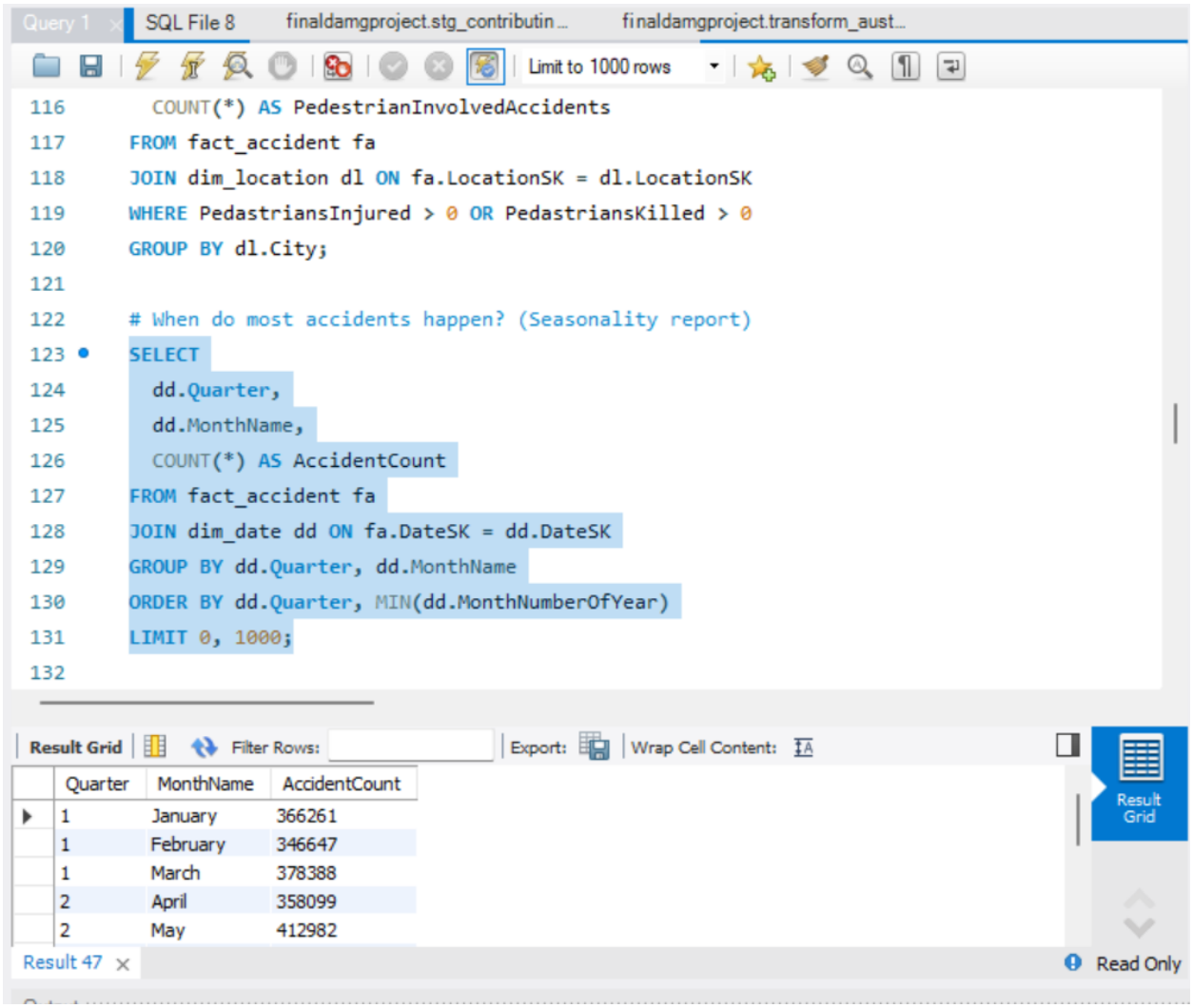
```
-- Overall
SELECT
  COUNT(*) AS PedestrianInvolvedAccidents
FROM fact_accident
WHERE PedastriansInjured > 0 OR PedastriansKilled > 0;

-- By City
SELECT
  dl.City,
  COUNT(*) AS PedestrianInvolvedAccidents
FROM fact_accident fa
JOIN dim_location dl ON fa.LocationSK = dl.LocationSK
WHERE PedastriansInjured > 0 OR PedastriansKilled > 0
GROUP BY dl.City;
```


Query 1
SQL File 8
finaldamgproject.stg_contributin ...
finaldamgproject.transform_aust...
Limit to 1000 rows
GROUP BY dl.City;
How often are pedestrians involved in accidents? (Overall and by city)
-- Overall
SELECT
COUNT(*) AS PedestrianInvolvedAccidents
FROM fact_accident
WHERE PedastriansInjured > 0 OR PedastriansKilled > 0;
-- By City
SELECT
dl.City,
COUNT(*) AS PedestrianInvolvedAccidents
FROM fact_accident fa
JOIN dim_location dl ON fa.LocationSK = dl.LocationSK
WHERE PedastriansInjured > 0 OR PedastriansKilled > 0
GROUP BY dl.City;
Result Grid
Filter Rows:
Export:
Wrap Cell Content:
PedestrianInvolvedAccidents
119320
Result Grid



5. When do most accidents happen? (Seasonality report)

```
SELECT
  dd.Quarter,
  dd.MonthName,
  COUNT(*) AS AccidentCount
FROM fact_accident fa
JOIN dim_date dd ON fa.DateSK = dd.DateSK
GROUP BY dd.Quarter, dd.MonthName
ORDER BY dd.Quarter, MIN(dd.MonthNumberOfYear)
LIMIT 0, 1000;
```



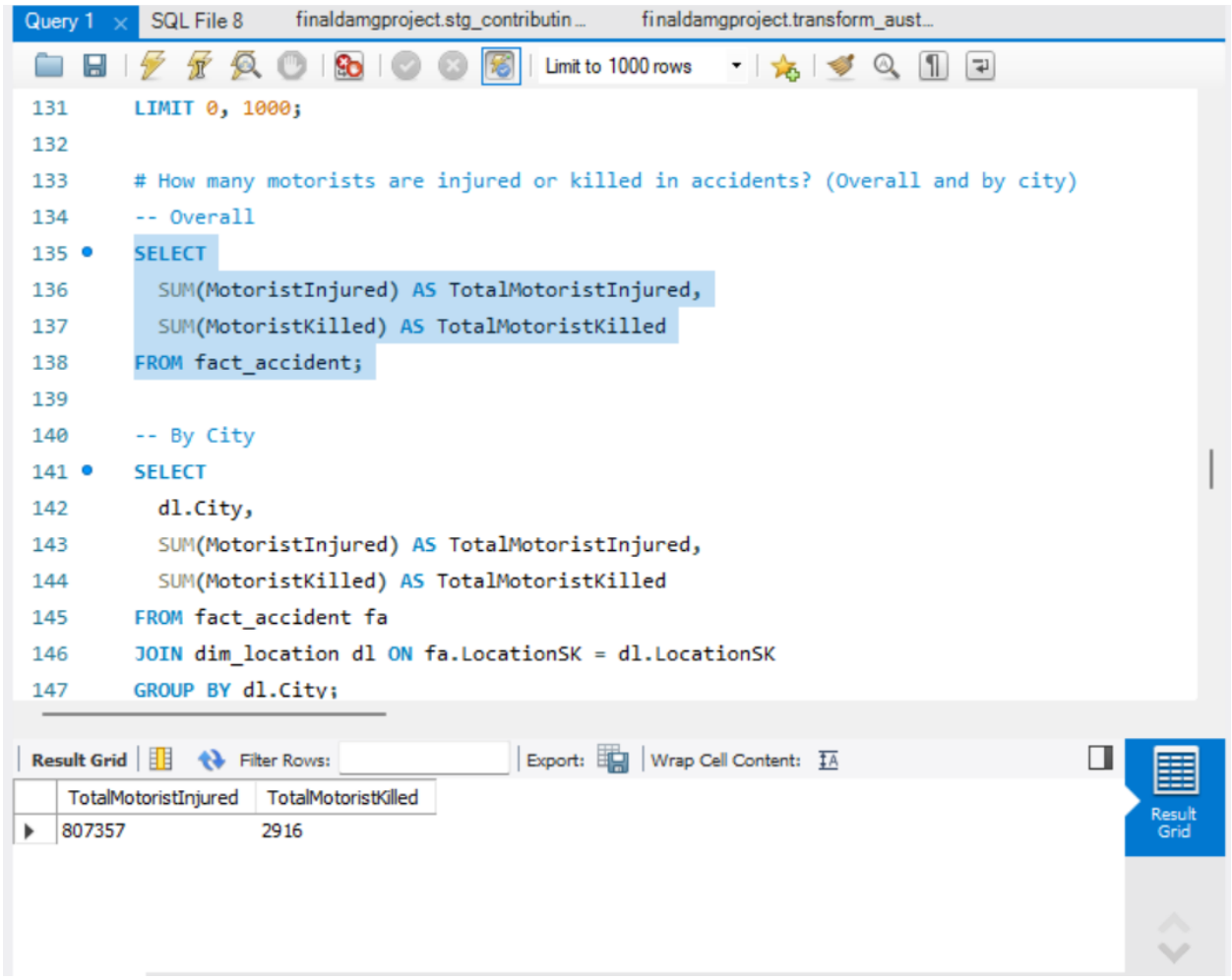
Quarter	MonthName	AccidentCount
1	January	366261
1	February	346647
1	March	378388
2	April	358099
2	May	412982

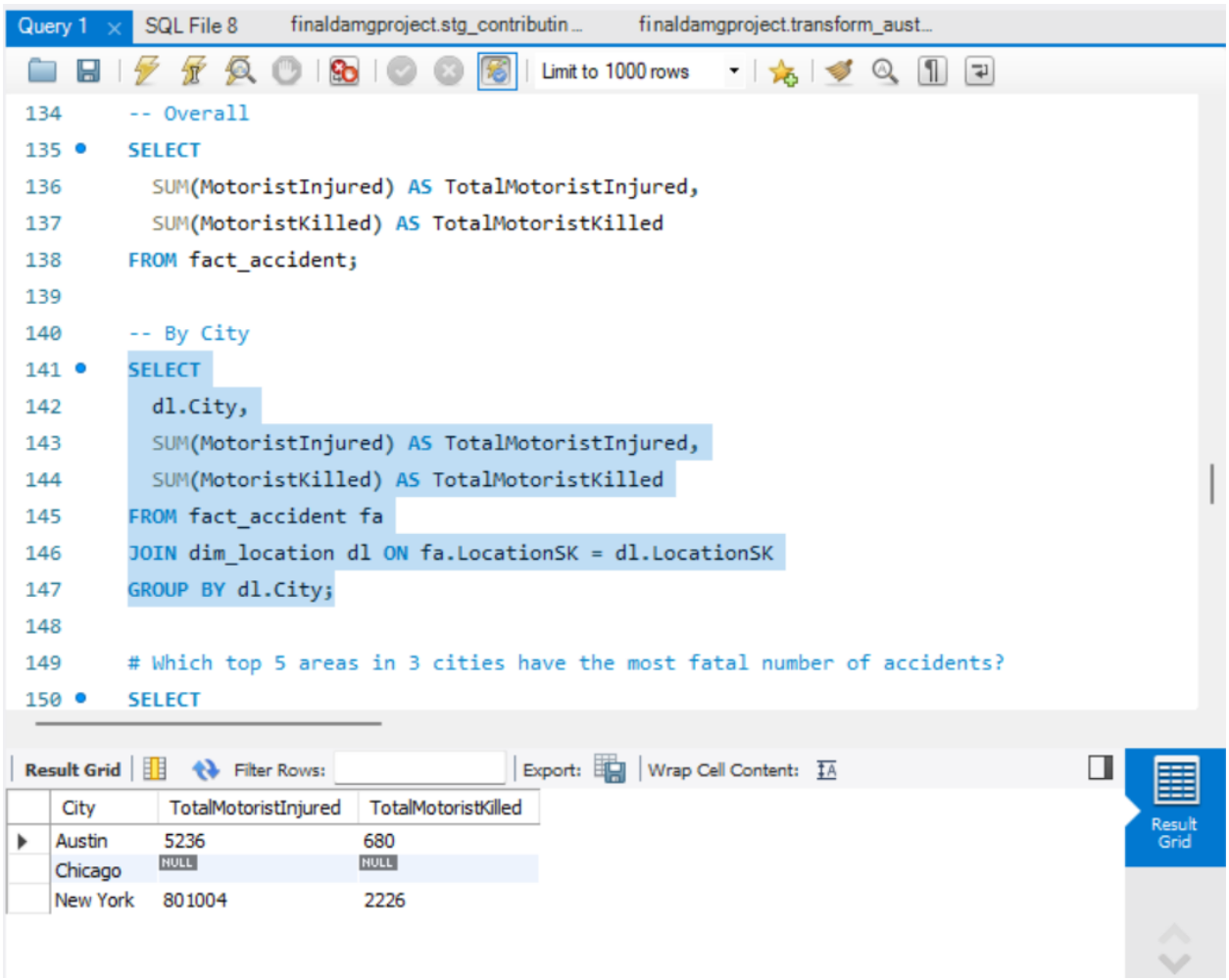
Result Grid | Filter Rows: | Expo

Quarter	MonthName	AccidentCount
1	January	366261
1	February	346647
1	March	378388
2	April	358099
2	May	412982
2	June	416479
3	July	408408
3	August	407148
3	September	411473
4	October	428588
4	November	397602
4	December	398222

6. How many motorists are injured or killed in accidents? (Overall and by city)

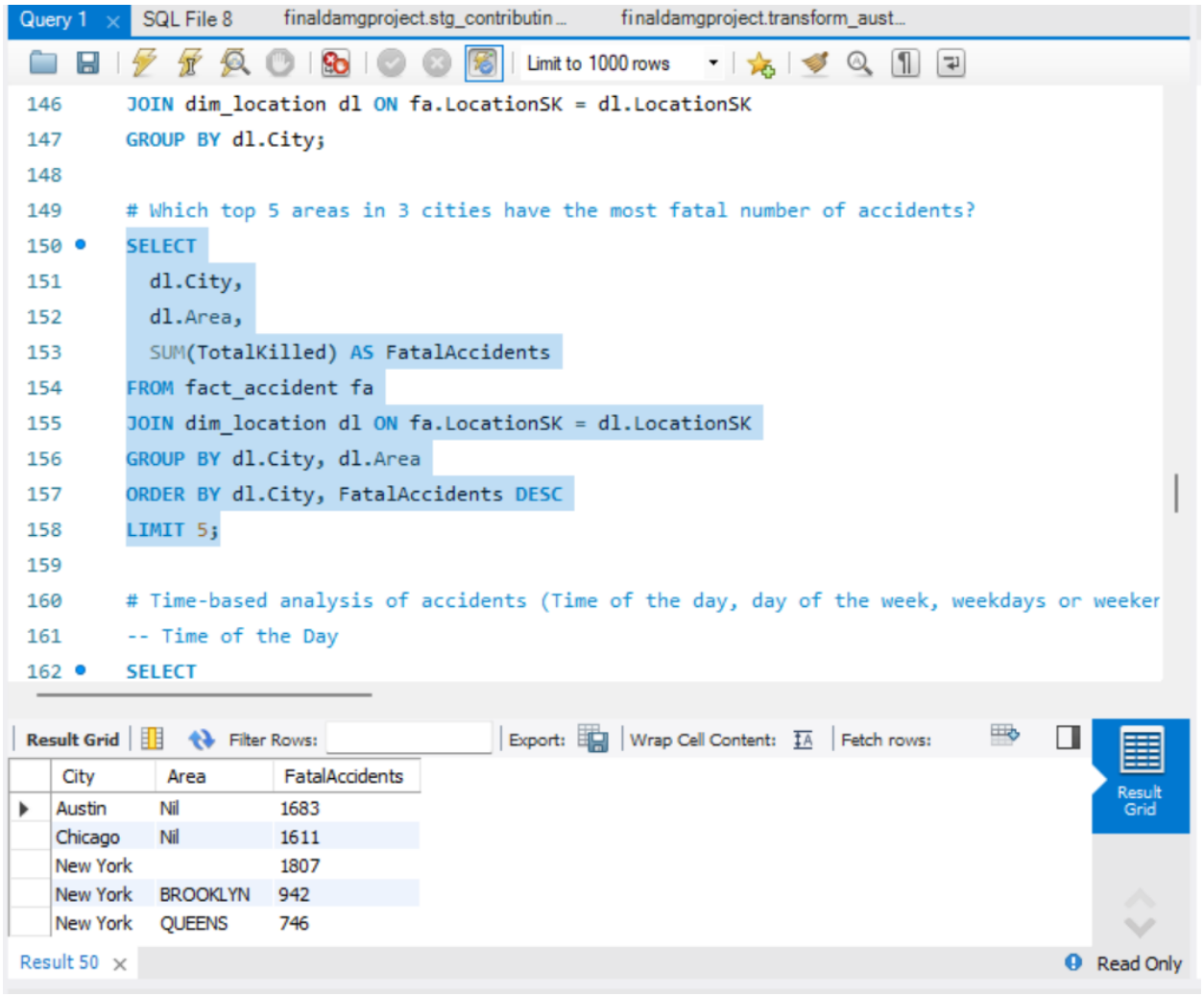
```sql
-- Overall
SELECT
  SUM(MotoristInjured) AS TotalMotoristInjured,
  SUM(MotoristKilled) AS TotalMotoristKilled
FROM fact_accident;

-- By City
SELECT
  dl.City,
  SUM(MotoristInjured) AS TotalMotoristInjured,
  SUM(MotoristKilled) AS TotalMotoristKilled
FROM fact_accident fa
JOIN dim_location dl ON fa.LocationSK = dl.LocationSK
GROUP BY dl.City;
```





7. Which top 5 areas in 3 cities have the most fatal number of accidents?

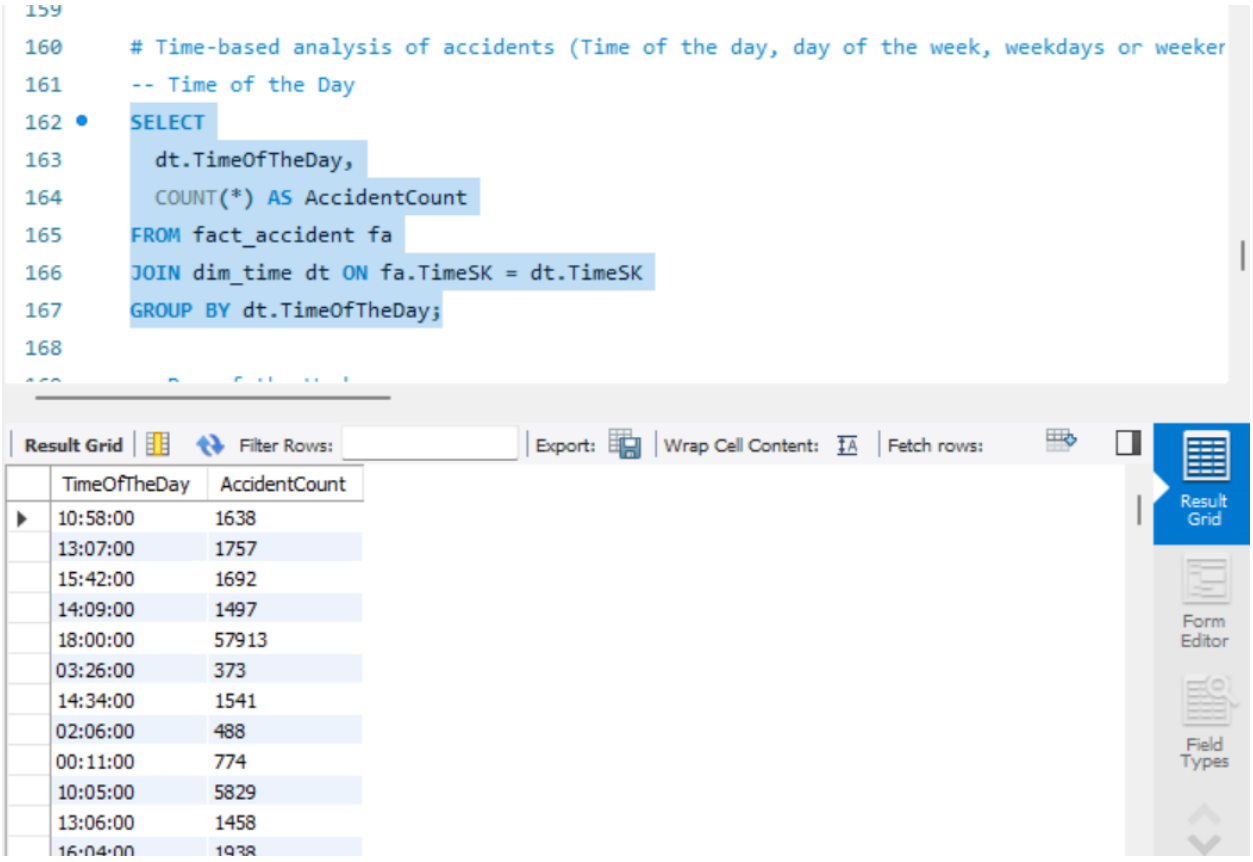
```
SELECT
  dl.City,
  dl.Area,
  SUM(TotalKilled) AS FatalAccidents
FROM fact_accident fa
JOIN dim_location dl ON fa.LocationSK = dl.LocationSK
GROUP BY dl.City, dl.Area
ORDER BY dl.City, FatalAccidents DESC
LIMIT 5;
```



8. Time-based analysis of accidents (Time of the day, day of the week, weekdays or weekends)

```
-- Time of the Day
SELECT
  dt.TimeOfTheDay,
  COUNT(*) AS AccidentCount
FROM fact_accident fa
JOIN dim_time dt ON fa.TimeSK = dt.TimeSK
GROUP BY dt.TimeOfTheDay;

-- Day of the Week
SELECT
  dd.DayNameOfWeek,
```

```
  COUNT(*) AS AccidentCount
FROM fact_accident fa
JOIN dim_date dd ON fa.DateSK = dd.DateSK
GROUP BY dd.DayNameOfWeek;

-- Weekdays vs Weekends
SELECT
  CASE
    WHEN dd.WeekendIndicator = 'Yes' THEN 'Weekend'
    ELSE 'Weekday'
  END AS DayType,
  COUNT(*) AS AccidentCount
FROM fact_accident fa
JOIN dim_date dd ON fa.DateSK = dd.DateSK
GROUP BY dd.WeekendIndicator;
```



```sql
-- Day of the Week
SELECT
  dd.DayNameOfWeek,
  COUNT(*) AS AccidentCount
FROM fact_accident fa
JOIN dim_date dd ON fa.DateSK = dd.DateSK
GROUP BY dd.DayNameOfWeek;

-- Weekdays vs Weekends
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

DayNameOfWeek	AccidentCount
Sunday	559067
Thursday	705153
Friday	765287
Wednesday	689591
Monday	668912
Tuesday	694456
Saturday	647831

Result Grid | Form Editor

Query 1 | SQL File 8 | finaldamgproject.stg_contributin... | finaldamgproject.transform_aust...

Limit to 1000 rows

```sql
-- Weekdays vs Weekends
SELECT
  CASE
    WHEN dd.WeekendIndicator = 'Yes' THEN 'Weekend'
    ELSE 'Weekday'
  END AS DayType,
  COUNT(*) AS AccidentCount
FROM fact_accident fa
JOIN dim_date dd ON fa.DateSK = dd.DateSK
GROUP BY dd.WeekendIndicator;
```

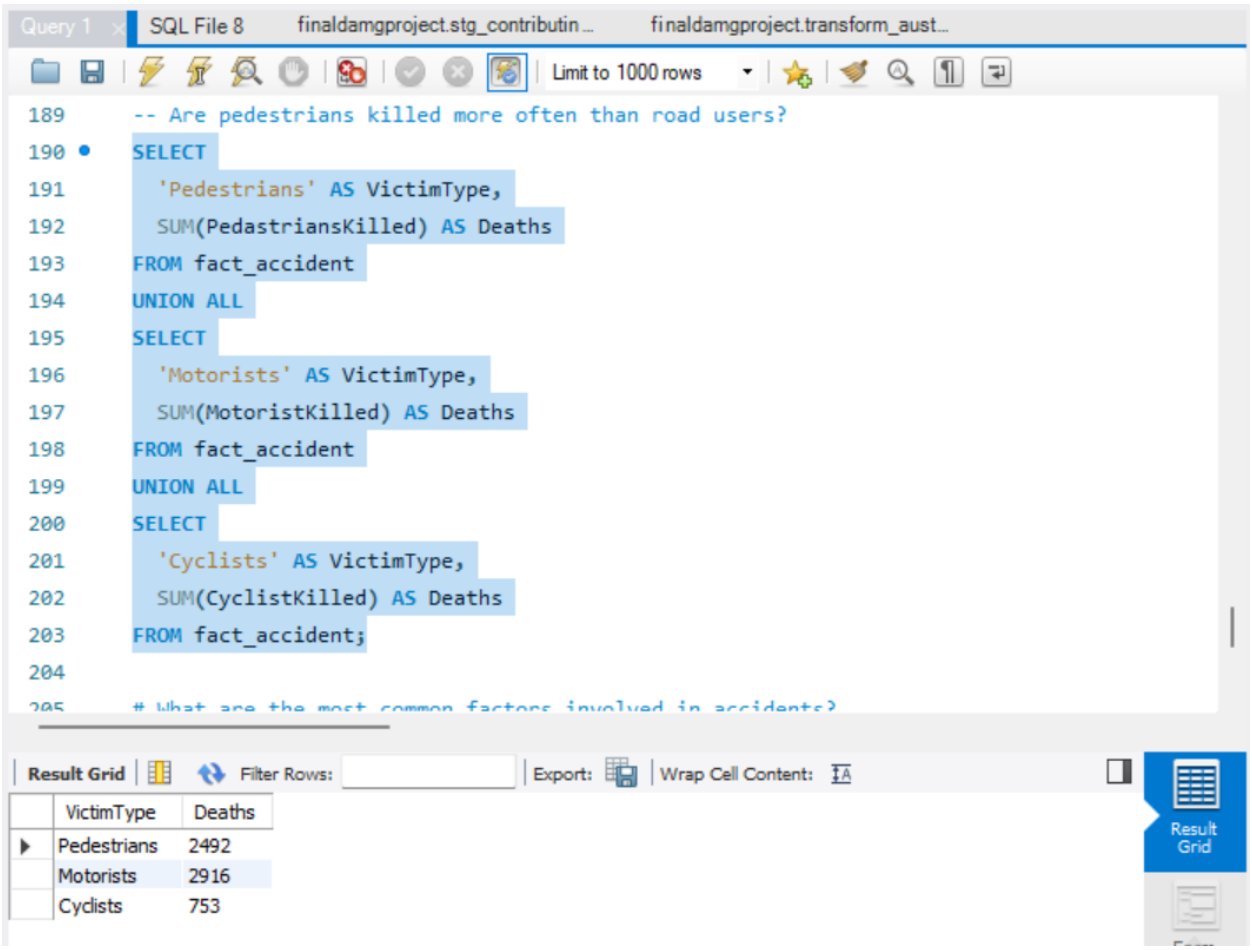
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

DayType	AccidentCount
Weekend	1206898
Weekday	3523399

Result Grid

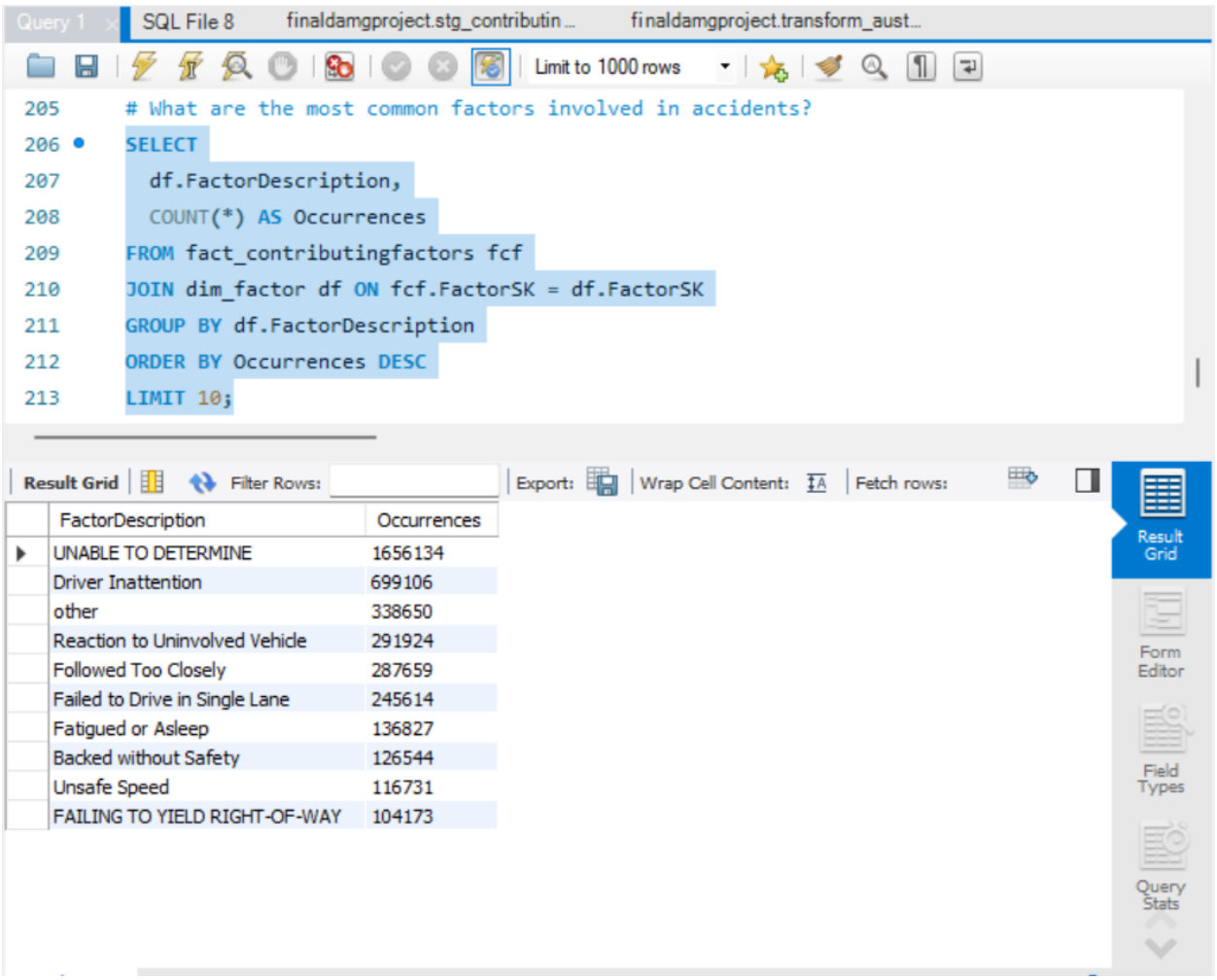
9. Fatality analysis

```
-- Are pedestrians killed more often than road users?
SELECT
  'Pedestrians' AS VictimType,
  SUM(PedastriansKilled) AS Deaths
FROM fact_accident
UNION ALL
SELECT
  'Motorists' AS VictimType,
  SUM(MotoristKilled) AS Deaths
FROM fact_accident
UNION ALL
SELECT
  'Cyclists' AS VictimType,
  SUM(CyclistKilled) AS Deaths
FROM fact_accident;
```



10. What are the most common factors involved in accidents?

```
SELECT
  df.FactorDescription,
  COUNT(*) AS Occurrences
FROM fact_contributingfactors fcf
JOIN dim_factor df ON fcf.FactorSK = df.FactorSK
GROUP BY df.FactorDescription
ORDER BY Occurrences DESC
LIMIT 10;
```


Query 1
SQL File 8
finaldamgproject.stg_contributin ...
finaldamgproject.transform_aust...
Limit to 1000 rows
205 # What are the most common factors involved in accidents?
206 SELECT
207 df.FactorDescription,
208 COUNT(*) AS Occurrences
209 FROM fact_contributingfactors fcf
210 JOIN dim_factor df ON fcf.FactorSK = df.FactorSK
211 GROUP BY df.FactorDescription
212 ORDER BY Occurrences DESC
213 LIMIT 10;
Result Grid
Filter Rows:
Export:
Wrap Cell Content:
Fetch rows:
FactorDescription
Occurrences
UNABLE TO DETERMINE 1656134
Driver Inattention 699106
other 338650
Reaction to Uninvolved Vehicle 291924
Followed Too Closely 287659
Failed to Drive in Single Lane 245614
Fatigued or Asleep 136827
Backed without Safety 126544
Unsafe Speed 116731
FAILING TO YIELD RIGHT-OF-WAY 104173
Result Grid
Form Editor
Field Types
Query Stats