

## Staging -

Job stg\_austin\_data 0.1

**Stage\_Austin\_Data**

Austin\_Data: 147750 rows in 12.19s, 12121.59 rows/s

tMap\_1: 147750 rows in 13.05s, 11320.1 rows/s

target\_ssms\_final1

Designer | Code

Job(stg\_austin\_data 0.1) | Context (stg\_austin\_data) | Component | Run (Job stg\_austin\_data) | Cloud Artifact

### Job stg\_austin\_data

**Basic Run**

Debug Run

Advanced settings

Target Exec

Memory Run

Execution

Run Kill Clear

Starting job stg\_austin\_data at 02:58 08/04/2024.  
[statistics] connecting to socket on port 3743  
[statistics] connected  
[statistics] disconnected

Job stg\_austin\_data ended at 02:58 08/04/2024. [Exit code = 0]

Default

Name	Value
Austin_Data_File	C:/Users/...
Austin_Data_H...	1
Austin_Data_Fi...	"\t"
Austin_Data_En...	UTF-8
Austin_Data_R...	"\n"
target_ssms_fin...	****
target_ssms_fin...	damg737

Job stg\_chicago\_data 0.1

Stage\_Chicago\_Data

Chicago\_Data

817723 rows in 43.01s  
19013.28 rows/s  
row1 (Main)

tMap\_1

817723 rows in 43.86s  
18645.64 rows/s  
chicago\_data (Main)

target\_ssms\_final1

DesignerCode

Job(stg\_chicago\_data 0.1)Context (stg\_chicago\_data)ComponentRun (Job stg\_chicago\_data)Cloud Artifact

Job stg\_chicago\_data

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

Execution

RunKillClear

Starting job stg\_chicago\_data at 03:00 08/04/2024.

[statistics] connecting to socket on port 3623

[statistics] connected

[statistics] disconnected

Job stg\_chicago\_data ended at 03:01 08/04/2024. [Exit code = 0]

Default

NameValue

Chicago\_Data...C:/Users//

Chicago\_Data...1

Chicago\_Data..."\t"

Chicago\_Data...UTF-8

Chicago\_Data..."\n"

target\_ssms\_fin...

target\_ssms\_fin...\*\*\*\*

target\_ssms\_fin...damg737

Job stg\_nyc\_data 0.1

Stage\_NYC\_Data

NYC\_Data

2075427 rows in 73.6s  
28199.12 rows/s  
row1 (Main order:1)

tMap\_1

2075427 rows in 74.21s  
27966.57 rows/s  
nyc\_data (Main)

target\_ssms\_final1

tLogRow\_1

1 rows in 73.76s  
row2 (Reject order:2)

Designer | Code

Job(stg\_nyc\_data 0.1) | Context (stg\_nyc\_data) | Component | Run (Job stg\_nyc\_data) | Cloud Artifact

**Job stg\_nyc\_data**

Basic Run  
Debug Run  
Advanced settings  
Target Exec  
Memory Run

Execution

Run Kill Clear

PARKING LOT).	0	0	0	0	0	0	0	0	Unspecified	null	null
---------------	---	---	---	---	---	---	---	---	-------------	------	------

[statistics] disconnected

Job stg\_nyc\_data ended at 03:03 08/04/2024. [Exit code = 0]

Default

Name	Value
NYC_Data_File	C:/Users//
NYC_Data_Hea...	1
NYC_Data_Fiel...	"\t"
NYC_Data_Enc...	UTF-8
NYC_Data_Row...	"\n"
target_ssms_fin...	
target_ssms_fin...	****
target_ssms_fin...	damg737

→ One row was rejected because it had “Unspecified” in an integer field.

Job staging2 0.1

147750 rows in 14.94s

90.22 rows/s

row1 (Main)

147750 rows in 15.52s

9519.97 rows/s

op\_austin (Main)

target\_ssms\_final1

tMap\_1

"austin\_stg"

2075427 rows in 110.61s

18636.59 rows/s

row2 (Main)

2075427 rows in 111.36s

18636.59 rows/s

op\_ny (Main)

target\_ssms\_final1

tMap\_2

"ny\_stg"

817723 rows in 66.2s

12270.02 rows/s

row3 (Main)

817723 rows in 66.64s

12270.02 rows/s

op\_chicago (Main)

target\_ssms\_final1

tMap\_3

"chicago\_stg"

Designer

Code

Job(staging2 0.1)

Context (staging2)

Component

Run (Job staging2)

Cloud Artifact

Job staging2

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

Execution

Run

Kill

Clear

Starting job staging2 at 03:05 08/04/2024.

[statistics] connecting to socket on port 3855

[statistics] connected

[statistics] disconnected

Job staging2 ended at 03:08 08/04/2024. [Exit code = 0]

Default

Name

Value

target\_final\_ss...

dbo

target\_final\_ss...

1433

target\_final\_ss...

target\_final\_ss...

damg737

target\_final\_ss...

\*\*\*\*

target\_final\_ss...

Project\_Fi

target\_final\_ss...

DESKTOP-

target\_ssms\_fin...

Project\_Fi



Job Austin\_CF 0.1Job Chicago\_CF 0.1Job NYC\_CF 0.1

```
graph LR; austin_stg -- "21537.9 rows/s" --> tMap_1; tMap_1 -- "537.9 rows/s" --> tMap_2; tMap_2 -- "7.9 rows/s" --> tNormalize_1; tNormalize_1 -- "22177.26 rows/s" --> tMap_3; tMap_3 -- "20729.8 rows/s" --> target_ssms_final1; Mapping_Data_ -- "85 rows in 0.03s" --> tMap_3;
```

DesignerCode

Job(Austin\_CF 0.1)Context (Austin\_CF)ComponentRun (Job Austin\_CF)Cloud Artifact

**Job Austin\_CF**

Basic RunDebug RunAdvanced settingsTarget ExecMemory Run

Execution

RunKillClear

Starting Job Austin\_CF at 03:12 08/04/2024.

[statistics] connecting to socket on port 3889

[statistics] connected

[statistics] disconnected

Job Austin\_CF ended at 03:12 08/04/2024. [Exit code = 0]

Default

Name	Value
target_ssms_fin...	damg737
target_ssms_fin...	1433
target_ssms_fin...	****
target_ssms_fin...	dbo
target_ssms_fin...	DESKTOP-
target_ssms_fin...	
target_ssms_fin...	Final_Proj
Mapping_Hea...	1

Job Chicago\_CF 0.1Job NYC\_CF 0.1

```
graph LR; chicago_stg["chicago_stg  
817723 rows in 41.76s  
49579.61 rows/s  
row1 (Main)"] --> tMap_1["tMap_1  
49577.74 rows/s  
combine_data (Main)"]; tMap_1 --> tMap_2["tMap_2  
49577.74 rows/s  
output1 (Main)"]; tMap_2 --> tNormalize_1["tNormalize_1  
777.74 rows/s  
row2 (Main)"]; tNormalize_1 --> tMap_3["tMap_3  
55.48 rows/s  
final_op (Main)"]; tMap_3 --> target_ssms_final1["target_ssms_final1  
88.26 rows/s"]; Mapping_Data_["Mapping_Data_  
85 rows in 0.03s  
2741.94 rows/s  
row3 (Main)"] --> tMap_4["tMap_4  
2741.94 rows/s  
trimmed (Main)"]; tMap_4 --> tNormalize_2["tNormalize_2  
2741.94 rows/s  
row4 (lookup)"]; tNormalize_2 --> target_ssms_final1;
```

DesignerCode

Job(Chicago\_CF 0.1)Context (Chicago\_CF)ComponentRun (Job Chicago\_CF)Cloud Artifact

### Job Chicago\_CF

**Basic Run**  
Debug Run  
Advanced settings  
Target Exec  
Memory Run

Execution

RunKillClear

Starting job Chicago\_CF at 03:12 08/04/2024.

[statistics] connecting to socket on port 3645

[statistics] connected

[statistics] disconnected

Job Chicago\_CF ended at 03:13 08/04/2024. [Exit code = 0]

Default

Name	Value
target_ssms_fin...	damg737
target_ssms_fin...	1433
target_ssms_fin...	****
target_ssms_fin...	dbo
target_ssms_fin...	DESKTOP
target_ssms_fin...	
target_ssms_fin...	Final_Proj
Mapping_Hea...	1

Job NYC\_CF 0.1

Designer | Code

Job(NYC\_CF 0.1) | Context (NYC\_CF) | Component | Run (Job NYC\_CF) | Cloud Artifact

### Job NYC\_CF

**Basic Run**

Execution

Run Kill Clear

Starting job NYC\_CF at 03:14 08/04/2024.  
 [statistics] connecting to socket on port 4049  
 [statistics] connected  
 [statistics] disconnected  
 Job NYC\_CF ended at 03:15 08/04/2024. [Exit code = 0]

Default

Name	Value
target_ssms_fin...	damg737
target_ssms_fin...	1433
target_ssms_fin...	****
target_ssms_fin...	dbo
target_ssms_fin...	DESKTOP
target_ssms_fin...	Final_Proj
Mapping_Hea...	1

DIM Loading →

1) Date →



The screenshot displays a data integration tool interface with a job flow diagram and a console window.

**Job Flow Diagram:**

- Inputs:**
  - "austin\_stg1" (Main): 147750 rows in 3.75s, 39389.5 rows/s
  - "ny\_stg2" (Main): 2075427 rows in 29.86s, 6956.9 rows/s
  - "chicago\_stg2" (Main): 817723 rows in 19.34s, 42286.7 rows/s
- Transformations:**
  - tMap\_1: 147750 rows in 52.94s, 2790.68 rows/s
  - tMap\_2: 2075427 rows in 52.94s, 39280.42 rows/s
  - tMap\_3: 817723 rows in 52.94s, 15445.06 rows/s
  - row2 (Main): 3040900 rows in 52.94s, 57416 rows/s
  - row3 (Main): 4287 rows in 53.73s, 79.79 rows/s
  - row4 (Main): 4287 rows in 53.73s, 79.79 rows/s
  - row5 (Uniques): 4287 rows in 53.73s, 79.79 rows/s
- Outputs:**
  - target\_ssms\_final1: 4287 rows in 53.73s, 79.79 rows/s

**Console Window:**

Job(date\_dim 0.1) Context(date\_dim) Component Run(Job date\_dim) Cloud Artifact

**Job date\_dim**

**Basic Run**

Execution

Run Kill Clear

Starting job date\_dim at 03:16 08/04/2024.  
 [statistics] connecting to socket on port 3763  
 [statistics] connected  
 [statistics] disconnected  
 Job date\_dim ended at 03:17 08/04/2024. [Exit code = 0]

**Default**

Name	Value
target_ssms_fin...	damg7370
target_ssms_fin...	Project_Final
target_ssms_fin...	1433
target_ssms_fin...	dbo
target_ssms_fin...	****
target_ssms_fin...	DESKTOP-F6...

2) Location →

Job location\_dim 0.1   Job time\_dim 0.1   Job source\_dim 0.1

Designer   Code

Job(location\_dim 0.1)   Context (location\_dim)   Component   Run (Job location\_dim)   Cloud Artifact

### Job location\_dim

**Basic Run**  
 Debug Run  
 Advanced settings  
 Target Exec  
 Memory Run

Execution

Run   Kill   Clear

Starting job location\_dim at 03:18 08/04/2024.  
 [statistics] connecting to socket on port 3876  
 [statistics] connected  
 [statistics] disconnected

Job location\_dim ended at 03:19 08/04/2024. [Exit code = 0]

Default

Name	Value
target_ssms_fin...	damg7370
target_ssms_fin...	Project_Final
target_ssms_fin...	1433
target_ssms_fin...	dbo
target_ssms_fin...	****
target_ssms_fin...	DESKTOP-F6...

3) Time →

Job time\_dim 0.1 x Job source\_dim 0.1

Designer Code

Job(time\_dim 0.1) Context (time\_dim) Component Run (Job time\_dim) Cloud Artifact

### Job time\_dim

**Basic Run**  
Debug Run  
Advanced settings  
Target Exec  
Memory Run

Execution

Run Kill Clear

Starting job time\_dim at 03:19 08/04/2024.  
[statistics] connecting to socket on port 3864  
[statistics] connected  
[statistics] disconnected

Job time\_dim ended at 03:21 08/04/2024. [Exit code = 0]

Default

Name	Value
target_ssms_fin...	damg7370
target_ssms_fin...	Project_Final
target_ssms_fin...	1433
target_ssms_fin...	dbo
target_ssms_fin...	****
target_ssms_fin...	DESKTOP-F6...

4) Source →

Job source\_dim 0.1
Job vehicles\_involved\_dim 0.1
Job Copy\_of\_Dim\_CF 0.1

```

graph LR
    austin_stg2["austin_stg2  
147750 rows in 7.76s  
19032.59 rows/s  
row1 (Main)"] --> tMap_1[tMap_1]
    ny_stg2["ny_stg2  
2075427 rows in 73.08s  
28397.83 rows/s  
row2 (Main)"] --> tMap_2[tMap_2]
    chicago_stg2["chicago_stg2  
817723 rows in 32.58s  
25099.7 rows/s  
row3 (Main)"] --> tMap_3[tMap_3]
    tMap_1 -- "147750 rows in 113.43s  
1302.67 rows/s  
out_austin (Merge order:1)" --> tUnite_1[tUnite_1]
    tMap_2 -- "2075427 rows in 113.43s  
18297.63 rows/s  
out_ny (Merge order:2)" --> tUnite_1
    tMap_3 -- "817723 rows in 113.43s  
7209.31 rows/s  
out_chicago (Merge order:3)" --> tUnite_1
    tUnite_1 -- "3040900 rows in 113.43s  
26788.13 rows/s  
row4 (Main)" --> tUniqRow_1[tUniqRow_1]
    tUniqRow_1 -- "3 rows in 114.68s  
0.03 rows/s  
rows (Uniques)" --> target_ssms_final1["target_ssms_final1"]

```

Designer
Code

Job(source\_dim 0.1)
Context (source\_dim)
Component
Run (Job source\_dim)
Cloud Artifact

### Job source\_dim

Basic Run
Debug Run
Advanced settings
Target Exec
Memory Run

Execution
Run
Kill
Clear

Starting job source\_dim at 03:22 08/04/2024.

[statistics] connecting to socket on port 3496

[statistics] connected

[statistics] disconnected

Job source\_dim ended at 03:24 08/04/2024. [Exit code = 0]

Default

Name	Value
target_ssms_fin...	damg7370
target_ssms_fin...	Project_Final
target_ssms_fin...	1433
target_ssms_fin...	dbo
target_ssms_fin...	****
target_ssms_fin...	DESKTOP-F6

5) Contributing\_Factor →

Job vehicles\_involved\_dim 0.1   Job Copy\_of\_Dim\_CF 0.1

Designer | Code

Job(Copy\_of\_Dim\_CF 0.1)   Context (Copy\_of\_Dim\_CF)   Component   Run (Job Copy\_of\_Dim\_CF)   Cloud Artifact

**Job Copy\_of\_Dim\_CF**

**Basic Run**

Debug Run   Advanced settings   Target Exec   Memory Run

Execution

Run   Kill   Clear

Starting job Copy\_of\_Dim\_CF at 03:25 08/04/2024.  
 [statistics] connecting to socket on port 3767  
 [statistics] connected  
 [statistics] disconnected

Job Copy\_of\_Dim\_CF ended at 03:26 08/04/2024. [Exit code = 0]

Name	Value
target_ssms_fin...	damg737
target_ssms_fin...	1433
target_ssms_fin...	****
target_ssms_fin...	dbo
target_ssms_fin...	DESKTOP
target_ssms_fin...	Final_Proj
target_ssms_fin...	Project_Fi

6) Vehicle\_Involved →

Job vehicles\_involved\_dim 0.1

Designer | Code

Job(vehicles\_involved\_dim 0.1) Context (vehicles\_involved\_dim) Component Run (Job vehicles\_involved\_dim) Cloud Artifact

### Job vehicles\_involved\_dim

**Basic Run**

Debug Run

Advanced settings

Target Exec

Memory Run

Execution

Run Kill Clear

Starting job vehicles\_involved\_dim at 03:26 08/04/2024.

[statistics] connecting to socket on port 3679

[statistics] connected

[statistics] disconnected

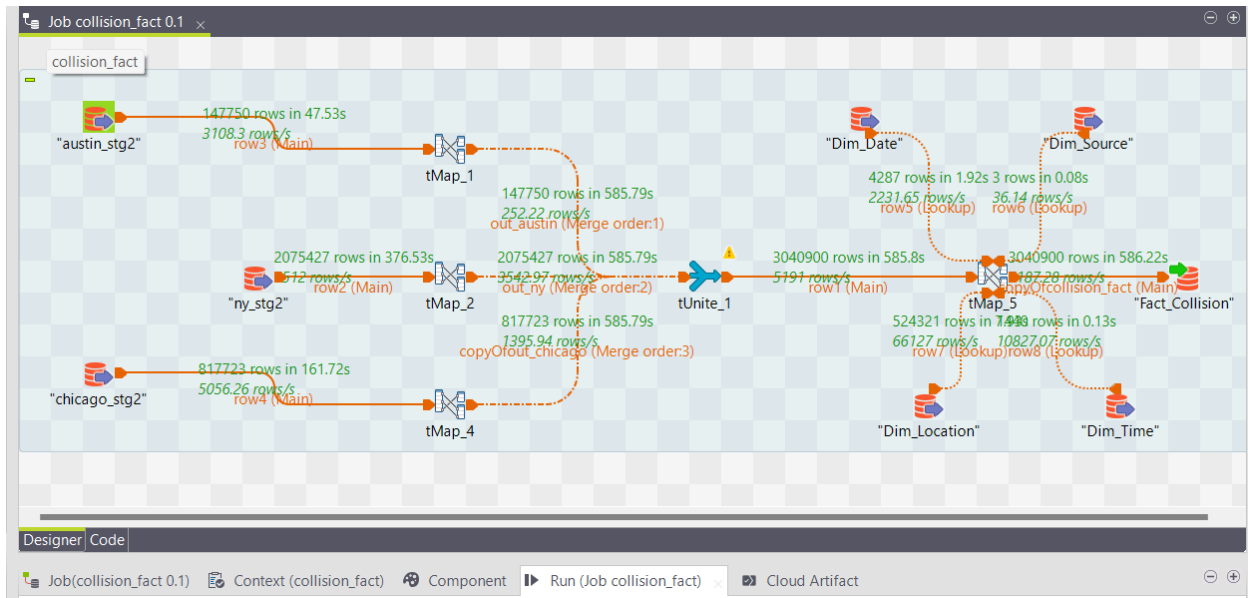
Job vehicles\_involved\_dim ended at 03:27 08/04/2024. [Exit code = 0]

Default

Name	Value
target_ssms_fin...	damg7370
target_ssms_fin...	Project_Final
target_ssms_fin...	1433
target_ssms_fin...	dbo
target_ssms_fin...	****
target_ssms_fin...	DESKTOP-F6...

Facts →

1) Collision →



→ One key was rejected due to the repetition.

2) Vehicle\_Involved →

Job vehicles\_involved\_fact 0.1

305032 rows in 44.98s  
"vehicles\_involved\_austin\_stg" tMap\_1  
305032 rows in 741.82s  
austin\_dim\_vehicles\_involved (Merge order:1)  
3924531 rows in 573.12s  
"vehicles\_involved\_ny\_stg" tMap\_3  
3924531 rows in 741.82s  
copyOfVeh1 (Merge order:2) row2 (Main) tUnite\_1  
5047286 rows in 741.82s  
5047286 rows in 742.89s  
copyOfVehicle\_type\_fact (Main) "Fact\_VehiclesInvolved"  
817723 rows in 123.73s  
"vehicles\_involved\_chicago\_stg" tMap\_6  
817723 rows in 741.82s  
chicago\_dim\_vehicles\_involved (Merge order:3)  
1855 rows in 1.49s  
1244.97 rows/s  
row5 (lookup)  
"Dim\_Vehicle\_Type"

Designer | Code

Job(vehicles\_involved\_fact 0.1) Context (vehicles\_involved\_fact) Component Run (Job vehicles\_involved\_fact) Cloud Artifact

**Job vehicles\_involved\_fact**

Basic Run  
Debug Run  
Advanced settings  
Target Exec  
Memory Run

Execution  
Run Kill Clear

Starting job vehicles\_involved\_fact at 03:55 08/04/2024.  
[statistics] connecting to socket on port 3842  
[statistics] connected  
[statistics] disconnected  
Job vehicles\_involved\_fact ended at 04:08 08/04/2024. [Exit code = 0]

Default

Name	Value
target_ssms_fin...	damg7370
target_ssms_fin...	Project_Final
target_ssms_fin...	1433
target_ssms_fin...	dbo
target_ssms_fin...	****
target_ssms_fin...	
target_ssms_fin...	DESKTOP-F6...

3) Factors →



Job contributing\_factor\_fact 0.1

Designer Code

Job(contributing\_factor\_fact 0.1) Context (contributing\_factor\_fact) Component Run (Job contributing\_factor\_fact) Cloud Artifact

### Job contributing\_factor\_fact

**Basic Run**

Debug Run

Advanced settings

Target Exec

Memory Run

**Execution**

Run Kill Clear

Starting job contributing\_factor\_fact at 03:39 08/04/2024.  
 [statistics] connecting to socket on port 3394  
 [statistics] connected  
 [statistics] disconnected

Job contributing\_factor\_fact ended at 03:54 08/04/2024. [Exit code = 0]

Default

Name	Value
target_ssms_fin...	damg7370
target_ssms_fin...	Project_Final
target_ssms_fin...	1433
target_ssms_fin...	dbo
target_ssms_fin...	****
target_ssms_fin...	DESKTOP-F6...

## Validation Queries →

1)

SQLQuery10.sql - D...al (damg7370 (51))\* SQLQuery9.sql - DE...al (damg7370 (72)) SQLQuery8.sql - DE...al (damg7370 (76))\* SQLQuery7.sql - DE...al (damg7370 (52))\*

```

select dt.time_of_day, count(dt.time_of_day)
from Fact_Collision f inner join Dim_Time dt
on f.TimeSK = dt.TimeSK
group by dt.time_of_day;

select dd.season, count(dd.season)
from Fact_Collision f inner join Dim_Date dd
on f.DateSK = dd.DateSK
group by dd.season;

select dd.day_of_the_week, count(dd.day_of_the_week)
from Fact_Collision f inner join Dim_Date dd
on f.DateSK = dd.DateSK
group by dd.day_of_the_week;

select cfd.contribution_description, count(cfd.contribution_description)
from Fact_Factors f inner join Dim_Contributing_factor cfd
on f.ContributingFactorSK = cfd.ContributingFactorSK

```

98 %

Results Messages

time_of_day	(No column name)
MORNING	618354
NIGHT	382649
DAWN	215812
AFTERNOON	749212
LATE NIGHT	445614
EVENING	629258

2)

```

select dd.season, count(dd.season)
from Fact_Collision f inner join Dim_Date dd
on f.DateSK = dd.DateSK
group by dd.season;

select dd.day_of_the_week, count(dd.day_of_the_week)
from Fact_Collision f inner join Dim_Date dd
on f.DateSK = dd.DateSK
group by dd.day_of_the_week;

select cfd.contribution_description, count(cfd.contribution_description)
from Fact_Factors f inner join Dim_Contributing_factor cfd
on f.ContributingFactorSK = cfd.ContributingFactorSK

```

98 %

Results Messages

season	(No column name)
Spring	716282
Summer	784599
Winter	733507
Fall	806511

day_of_the_week	(No column name)
1	368737
4	440445
3	443249
5	449692
2	429002
6	487835
7	421939

Query executed successfully. DESKTOP-F69117S (16.0 RTM) damg7370 (51) Project\_Final 00:00:00 11 rows

3)

```

select cfd.contribution_description, count(cfd.contribution_description)
from Fact_Factors f inner join Dim_Contributing_factor cfd
on f.ContributingFactorSK = cfd.ContributingFactorSK
group by cfd.contribution_description

select avg(speed_limit) as avg_speed_limit, max(speed_limit) as max_speed_limit, min(speed_limit) as min_speed_limit
from Fact_Collision
where speed_limit >= 0;

select vtd.vehicle_type, count(vtd.vehicle_type)

```

%

Results Messages

contribution_description	(No column name)
Fatigued or Asleep	91778
Distraction in Vehicle	10348
Failed to Yield ROW - To Pedestrian	87
Wrong Side - Not Passing	6090
Turned when Unsafe	35761
Overtake and Pass Insufficient Clearance	52824
Wrong Side - Approach or Intersection	10
Failed to Drive in Single Lane	185702
Failed to Yield ROW - Turning Left	636
Failed to Yield ROW - Yield Sign	522
Cell/Mobile Device Use - Texting	566
UNABLE TO DETERMINE	620860
Failed to Yield ROW - Open Intersection	189
Drove Without Headlights	272

avg_speed_limit	max_speed_limit	min_speed_limit
30	99	0

4)

SQLQuery10.sql - D...al (damg7370 (51))\* SQLQuery9.sql - DE...al (damg7370 (72)) SQLQuery8.sql - DE...al (damg7370 (76))\* SQLQuery7.sql - DE...al (damg7370 (52))\*

```
from Fact_Collision
where speed_limit >= 0;

select vtd.vehicle_type, count(vtd.vehicle_type)
from Fact_VehiclesInvolved f inner join Dim_Vehicle_Type vtd
on f.VehicleTypeSK = vtd.VehicleTypeSK
group by vehicle_type;
```

98 %

Results Messages

	vehicle_type	(No column name)
1	school bus	457
2	moped	4908
3	tow truck / wrecker	2719
4	garba	71
5	cat	1
6	ltrl	2
7	self	42
8	revel scoo	5
9	self insur	17
10	stree	63
11	a	4
12	50 cc moto	1
13	transit mt	1
14	utility	37
15	fork lift	44
16	mcy	7
17	geico	2
18	mini van	5
19	delivery t	40
20	ems truck	7
21	work truck	2
22	unmarked v	1

Query executed successfully.

DESKTOP-F69117S (16.0 RTM) | damg7370 (51) | Project\_Final | 00:00:00 | 1,855 rows