Final Project Dimensional Model

CIS 9440 - Data Warehousing and Analytics
Project title: NYC Motor Vehicle Collision Transparency Data Warehouse Project
Final Project Milestone 2
Group Number: 5
Student(s): Gabriel Fernandez, Jason Jiang

For Milestone #2, you/your project group will use Dimensional Modeling to turn your >= 1 data sources into Fact Tables and Dimensions. To do so, **submit each of the following as a PDF.**
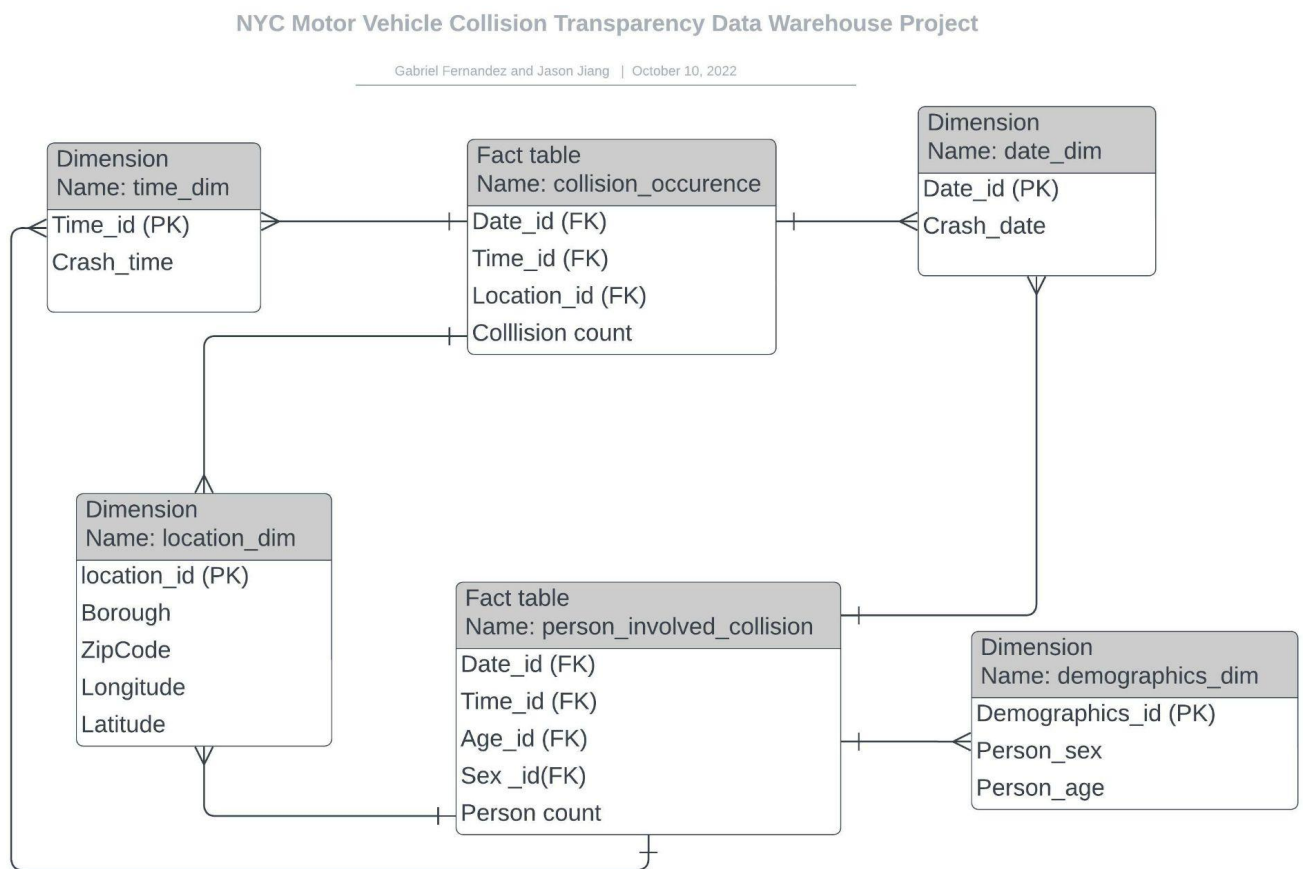Please include your group number and members of your Final Project team as a "**Text Submission**" when you submit the Milestone. All members from each Project Group must submit the Milestone.

(1) Complete a Kimball Bus Matrix for your project. This Matrix must include all your Fact tables on the left-hand side as rows (this project requires 1 Fact Table, but may have more than 1), and all your Dimensions on the top as columns. Use an 'X' to show which Fact Tables will be connected to each Dimension. Dimensions that are connected to more than 1 Fact Table are "Conformed Dimensions". The link to the Bus Matrix template to complete and submit as a PDF is here: link.

Dimensions

Business Processes (Facts)	Date	Time	Location	Demographics	
collision_occurance	x	x	x		
person_involved_collision	x	x	x	x	

Link: 🗓 Group5_Kimball BUS Matrix

(2) Submit a Dimensional Model that includes all your Fact Tables and Dimensions. This model will be either a Star Schema or a Multi Fact Star Schema (or "Constellation Schema"). The Dimensional Model should include Primary Keys and Attributes in each Dimension, and Foreign Keys and Measures in each Fact Table. You may use any medium necessary to create this model, such as draw.io, lucidchart.com, or copying and editing the template we used in class. You may use any medium necessary to create this model, such as draw.io, lucidchart.com, or copying and editing the template we used in class (link).

NYC Motor Vehicle Collision Transparency Data Warehouse Project

Gabriel Fernandez and Jason Jiang | October 10, 2022

Dimension
Name: time_dim
Time_id (PK)
Crash_time

Fact table
Name: collision_occurence
Date_id (FK)
Time_id (FK)
Location_id (FK)
Colllision count

Dimension
Name: date_dim
Date_id (PK)
Crash_date

Dimension
Name: location_dim
location_id (PK)
Borough
ZipCode
Longitude
Latitude

Fact table
Name: person_involved_collision
Date_id (FK)
Time_id (FK)
Age_id (FK)
Sex _id(FK)
Person count

Dimension
Name: demographics_dim
Demographics_id (PK)
Person_sex
Person_age

Link to Lucidchart: NYC Motor Vehicle Collision Transparency Data Warehouse Project

(3) Now that you have a Dimensional Model, refine and submit a list of your **5 best project KPIs**. Your KPIs should guide your Data Warehouse users/audience to make data-driven decisions. The best KPIs are not merely nice to know, they may be acted upon. Along with your KPI's, write a short description (>= 2 sentences per KPI) of how you envision your Data Warehouse users/audience to use the KPI's. Remember, your Data Warehouse audience may be hypothetical.

High-level Business Initiative:
We want to discover where and when most motor vehicle accidents happen in NYC and which demographics are more affected. Our group ultimately wants to bring transparency to the traffic collisions in NYC.

 5 best project KPIs

1. Number of collisions per hour

   This KPI can show which time collisions are more likely to occur. The Metropolitan Transportation Authority can use this insight to increase public transportation during crash- prone hours to disincentivize driving. Also, the New York City Department of Transportation can use this KPI to evaluate the proposed congestion pricing initiative. If congestion pricing reduces traffic congestion, this could translate into reduced collisions since drivers will have a better commute experience.

2. Number collisions by Zip Code

   This KPI can show the zip codes with the most collisions. The New York City Department of Transportation can prioritize the zip codes with higher collisions with initiatives to reduce and prevent collisions. Also, they can compare successful interventions in the boroughs with fewer collisions and allocate more funds to the more affected boroughs.

3. Number of collisions by Age
   PPublic safety officials can create educational campaigns targeting the age groups more likely to experience a collision. Influencers for each age group can persuade people to behave in ways that reduce traffic collisions.

4. Number of collisions by Sex
   Public safety officials can create educational campaigns customized for each sex to reduce traffic collisions.

5. Number of collisions by Date
   This KPI allows us to investigate which date collisions are more likely to occur and can help identify any seasonalities (patterns) that exist.