



# Traffic Crashes in Chicago

Team 3: Hoyu Li, Chen Liu, Siran Fang, Yansiyu Pan

Database Design and Implementation

# AGENDA

---

- Summary
- Research Objectives
- Solution Overview
- ETL/Database Design
- Analysis & Visualization
- Conclusion



# Executive Summary

---

Our project aims to investigate and improve the safety of commute in Chicago. The project uses traffic crash data as collected by authorities across Chicago, including the Chicago Police Department, to compile and analyze trends in crashes, injuries and fatalities.

This project examines traffic crashes in Chicago in 2019 and aims to provide insights regarding the contributing factors for the majority of road accidents. Tools such as MySQL and Tableau were used to analyze and visualize data. Analysis was done from multiple perspectives, including weather and lighting conditions, timing, drivers' demographics, vehicle conditions etc.



# Traffic Crash Report

The Traffic Crash Report (Form SR1050 in Illinois) prepared by a law enforcement officer

ILLINOIS TRAFFIC CRASH REPORT												Sheet _____ of _____ Sheets		* XXXXXXXX *									
BRAG	PEEV	SPDO	TRNG	WEAT	DRDA	VIS	VEHD	LONT	COLL	MANY	FPA	FPL				ACTVY CRASH REPORT NO.	TRAVN						
UH	U2			U1	U2	U1	U2	U1	U2	U1	U2												
INVESTIGATING AGENCY												DAMAGE TO VEHICLE ONE PERSON'S VEHICLE / PROPERTY	\$500 OR LESS \$501 - \$1,000 OVER \$1,000	TYPE OF REPORT ON SCENE NOT ON SCENE (DESK REPORT) AMENDED	<input type="checkbox"/> A. No Injury / Drive Away <input type="checkbox"/> B. Injury and / or Tow Due To Crash	DATE OF CRASH CIRCLE DAY OF WEEK SU MO TU WE TH FR SA	TIME AM PM	LARS CODE	VEHT				
ADDRESS NO.				HIGHWAY or STREET NAME								City	Township	INTERSECTION RELATED <input type="checkbox"/> Yes <input type="checkbox"/> No	DATE OF CRASH CIRCLE DAY OF WEEK SU MO TU WE TH FR SA	NUMBER MOTOR VEHICLES INVOLVED	LARS CODE	U1					
BLOCK FT / MI N E S W				AT INTERSECTION WITH (NAME OF INTERSECTION OR ROAD FEATURE)								COUNTY	PRIVATE PROPERTY <input type="checkbox"/> Yes <input type="checkbox"/> No	HIT & RUN <input type="checkbox"/> Yes <input type="checkbox"/> No			NO LANES						
NAME <input type="checkbox"/> DRIVER <input type="checkbox"/> PASSED - NO DRIVER <input type="checkbox"/> PED <input type="checkbox"/> PEDAL <input type="checkbox"/> EQUIP <input type="checkbox"/> NMV <input type="checkbox"/> NOV												DATE OF BIRTH mo / day / yr	MAKE	MODEL	YEAR	FRONT	Y N	U2					
LAST FIRST M.I. STREET ADDRESS												SEX MALE FEMALE	PLATE NO.	STATE	YEAR	TOPED TOWED FIRED FIRE	ALIGN	RESUR					
CITY STATE ZIP				INJURY EJECT								VIN					VEHNU						
TELEPHONE				DRIVER LICENSE NO. STATE CLASS								VEHICLE OWNER (LAST, FIRST M.I.)	INSURANCE CO.			U1							
TANDEM TO				EMS AGENCY								OWNER ADDRESS (STREET, CITY, STATE, ZIP)	TELEPHONE	POLICY NO.			U2						
NAME <input type="checkbox"/> DRIVER <input type="checkbox"/> PASSED - NO DRIVER <input type="checkbox"/> PED <input type="checkbox"/> PEDAL <input type="checkbox"/> EQUIP <input type="checkbox"/> NMV												DATE OF BIRTH mo / day / yr	MAKE	MODEL	YEAR	FRONT	Y N	U2					
LAST FIRST M.I. STREET ADDRESS												SEX MALE FEMALE	PLATE NO.	STATE	YEAR	TOPED TOWED FIRED FIRE	ALIGN	RESUR					
CITY STATE ZIP				INJURY EJECT								VIN					VEHNU						
TELEPHONE				DRIVER LICENSE NO. STATE CLASS								VEHICLE OWNER (LAST, FIRST M.I.)	INSURANCE CO.			U1							
TANDEM TO				EMS AGENCY								OWNER ADDRESS (STREET, CITY, STATE, ZIP)	TELEPHONE	POLICY NO.			U2						
LURD (SEAT)				RSRD (SEAT)								PASSENGERS & WITNESSES ONLY (NAME / ADDRESS / TEL)				RSRD	ITEMS			U1			
/ /																				U2			
/ /																				U1			
/ /																				U2			
/ /																				U1			
/ /																				U2			
/ /																				U1			
/ /																				U2			
UNIT	EVNO	INJNT	EVNT	LOC	DAMAGED PROPERTY OWNER NAME								DAMAGED PROPERTY				CONTRIBUTORY CAUSE(S)		POSTED SPEED LIMIT	* P 0 1 0 9 *			
1	<input type="checkbox"/>				PROPERTY OWNER ADDRESS								CITY STATE ZIP				PRIMARY						
2	<input type="checkbox"/>				ARREST NAME								SECTION				SECONDARY						
3	<input type="checkbox"/>				ARREST NAME								SECTION				CITATION NO.						
UNIT	1	<input type="checkbox"/>			OFFICER ID.								BEAT / DIST.				SUPERVISOR ID.		DATE POLICE NOTIFIED mm / dd / yy		TIME NOTIFIED AM PM		
2	<input type="checkbox"/>				SIGNATURE														COUNT DATE mm / dd / yy		COUNT TIME AM PM		
3	<input type="checkbox"/>																						
REMEMBER TO USE BLACK INK, PRESS HARD, PRINT LEGIBLY AND COMPLETE ALL REQUIRED FIELDS!																							

\*IF YES TO HAZMAT SPILL OR COM VEH, COMPLETE COMMERCIAL MOTOR VEHICLE AREA ON BACK.

# Data Sources

## Chicago Data Portal - Transportation

- Traffic Crashes
- Traffic Crashes - Vehicles
- Traffic Crashes - People

Link:

<https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if>

Project data span: 01/01/2019 -  
10/31/2019

The screenshot shows the Chicago Data Portal interface. At the top, there is a navigation bar with various categories. Below it, the specific dataset 'Traffic Crashes - Crashes' is highlighted. The main content area displays the dataset's purpose and some descriptive text. At the bottom, there are links to other related datasets.

RD\_NO CRASH\_DATE\_EST\_I CRASH\_DATE POSTED\_SPEED\_LIMIT TRAFFIC\_CONTROL\_DEVICE DEVICE\_CONDITION WEATHER\_CONDITION LIGHTING\_CONDITION

GCTC CITY OF CHICAGO INCORPORATED 4th March 1837 CHICAGO DATA PORTAL

## Traffic Crashes - Crashes

Transportation

Crash data shows information about each traffic crash on city streets within Chicago limits and under the jurisdiction of Chicago Police Department (CPD). Data is shown as is from the electronic crash reporting system (E-Crash) at CPD, excluding personally identifiable information. Records are added to the data portal via the E-Crash system.

[More](#)

### Featured Content Using this Data

[Traffic Crashes - Vehicles](#) [Traffic Crashes - People](#)

# Research Objectives

---

- Analyze the number of crashes based on the following parameters: timing, weather, lighting, sex of drivers, location, and vehicle conditions
- Traffic Management Authority can better allocate personnel and resources to handle emergency situations
- Citizens can better plan their travel to avoid dangerous driving conditions and traffic jam

## Business Use Case

City Leadership - Public Resource Planning

Prioritize the street segments and intersections that need safety improvement

Traffic Signals Modernization /Intersection Improvement/Street Widening/Enforcement Efforts/Roadway Safety Audit

Industry Purposes - Planning agencies, automobile manufacturers, insurance companies, media, legislatures all use crash data for planning and operational agenda

# Tools For Data Processing

---

Data Source



CHICAGO  
DATA PORTAL

ETL Process



Data Storage



Google Cloud Platform

Data Modeling  
& Analysis



Data  
Visualization &  
Dashboard



Insights &  
Reporting



# Solution Overview

## Datasets

Crashes Dataset

People Dataset

Vehicle Dataset

Location Dataset

Primary Contributor  
Dataset



## Data Collectors

Data Access APIs

CSV Extracts

Processing Scripts



## Relational Database

SQL for manipulation  
of data for storage

Queries for data  
visualizations

Tables for data sources  
and analysis results



## Analytics & Visualizations

Visualizations of trends  
and patterns



# Traffic Crash

- Crash data details information of each traffic crash on city streets within the City of Chicago limits.
- The original dataset contains 48 columns. After excluding irrelevant variables, we retain 33 columns.

## Screenshots of Crash Variables

RD_NO	CRASH_DATE	POSTED_SPEED_LIMIT	TRAFFIC_CONTROL_DEVICE	DEVICE_CONDITION	WEATHER_CONDITION	LIGHTING_CONDITION	FIRST_CRASH_TYPE	TRAFFICWAY_TYPE
1. JC334993	2019-07-04T00:00:00Z	45	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD	REAR END	DIVIDED - W/MEDIAN BARRIER
2. JC370822	2019-07-30T00:00:00Z	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT	TURNING	DIVIDED - W/MEDIAN (NOT RAISED)
MOST_SEVERE_INJURY	INJURIES_TOTAL	INJURIES_FATAL	INJURIES_INCAPACITATING	INJURIES_NON_INCAPACITATING	INJURIES_REPORTED_NOT_EVIDENT	INJURIES_NO_INDICATION	INJURIES_UNKNOWN	

ALIGNMENT	ROADWAY_SURFACE_COND	CRASH_TYPE	DAMAGE	PRIM_CONTRIBUTORY_CAUSE	SEC_CONTRIBUTORY_CAUSE	STREET_NO	STREET_DIRECTION	STREET_NAME	NUM_UNITS
STRAIGHT AND LEVEL	DRY	NO INJURY / DRIVE AWAY	OVER \$1,500	1	NOT APPLICABLE	300	N	LAKE SHORE DR SB	
STRAIGHT AND LEVEL	DRY	NO INJURY / DRIVE AWAY	OVER \$1,500	2	IMPROPER TURNING/NO SIGNAL	8201	S	DR MARTIN LUTHER KING JR DR	
STRAIGHT AND LEVEL	DRY	NO INJURY / DRIVE AWAY	\$501 - \$1,500	3	NOT APPLICABLE	6747	S	CREGIER AVE	1
STRAIGHT AND LEVEL	DRY	NO INJURY / DRIVE AWAY	\$501 - \$1,500	4	NOT APPLICABLE	554	N	FRANKLIN ST	1
STRAIGHT AND LEVEL	DRY	NO INJURY / DRIVE AWAY	\$501 - \$1,500	4	UNABLE TO DETERMINE	3700	N	WESTERN AVE	1

## Important Variables

Prim\_Contributory

Crash\_day\_of\_week

Damage

Crash\_Hour

Weather\_Condition

Crash\_Month

Lighting\_Condition

Most\_Severe\_Injury

## Screenshots of People Variables

SEX	AGE	DRIVERS_LICENSE_STATE	DRIVERS_LICENSE_CLASS	INJURY_CLASSIFICATION	DRIVER_ACTION
M				NO INDICATION OF INJURY	FOLLOWED TOO CLOSELY
M	47	NC		NO INDICATION OF INJURY	NONE
PERSON_ID	PERSON_TYPE	RD_NO	VEHICLE_ID	CRASH_DATE_1	CITY
1. 0555704	DRIVER	JC101289	530296	2019-01-02T00:00:00Z	
2. 0555705	DRIVER	JC101289	530292	2019-01-02T00:00:00Z	NEWPORT
3. 0555706	DRIVER	JC101293	530302	2019-01-02T00:00:00Z	
4. 0555709	DRIVER	JC101298	530295	2019-01-02T00:00:00Z	
5. 0555714	DRIVER	JC101323	530304	2019-01-02T00:00:00Z	CHICAGO
6. 0555715	DRIVER	JC101324	530305	2019-01-02T00:00:00Z	CHICAGO
7. 0555717	DRIVER	JC101337	530307	2019-01-02T00:00:00Z	

# Traffic Crash - People

- This data contains information about people involved in a crash and if any injuries were sustained.
- Each record corresponds to an occupant in a vehicle listed in the Crash dataset.

### Important Variables

SEX

AGE

PERSON\_TYPE

INJURY\_CLASSIFICATION

## Screenshots of Vehicle Variables

CRASH_UNIT_ID	RD_NO	CRASH_DATE_1	UNIT_NO	UNIT_TYPE	VEHICLE_ID	MAKE	MODEL	LIC_PLATE_STATE	VEHICLE_YEAR	VEHICLE_DEFECT
1. 555704	JC101289	2019-01-02T00:00:00Z	1	DRIVER	530296	UNKNOWN	UNKNOWN	IL		NONE
2. 555705	JC101289	2019-01-02T00:00:00Z	2	DRIVER	530292	VOLVO	S80	NC	2009	UNKNOWN
3. 555706	JC101293	2019-01-02T00:00:00Z	1	DRIVER	530302	UNKNOWN	UNKNOWN	IL		UNKNOWN
4. 555707	JC101293	2019-01-02T00:00:00Z	2	PARKED	530297	NISSAN	NISSAN MAXIMA	IL	2013	UNKNOWN
5. 555709	JC101298	2019-01-	1	DRIVER	530295	DODGE	CARAVAN	IL	2003	UNKNOWN
VEHICLE_TYPE	VEHICLE_USE	TRAVEL_DIRECTION	MANEUVER		DODGE	TOWED_I	FIRE_I	OCCUPANT_CNT	EXCEED_SPEED_LIMIT_I	FIRST_CONTACT_POINT
PASSENGER	UNKNOWN/NA	S	STRAIGHT AHEAD					1		FRONT
PASSENGER	PERSONAL	S	STRAIGHT AHEAD					1		REAR
UNKNOWN/NA	UNKNOWN/NA	N	UNKNOWN/NA					1		UNKNOWN
PASSENGER	PERSONAL	N	PARKED		Y			0		REAR-LEFT
VAN/MINI-VAN	UNKNOWN/NA	N	UNKNOWN/NA		N			1		FRONT
SPORT UTILITY VEHICLE (SUV)	PERSONAL	N	AVOIDING VEHICLES/OBJECTS		Y			1		FRONT
PASSENGER	PERSONAL	W	STRAIGHT AHEAD					1		FRONT-LEFT
PASSENGER	PERSONAL	W	PARKED					0		REAR-RIGHT

### Important Variables

UNIT\_TYPE

MAKE

MODEL

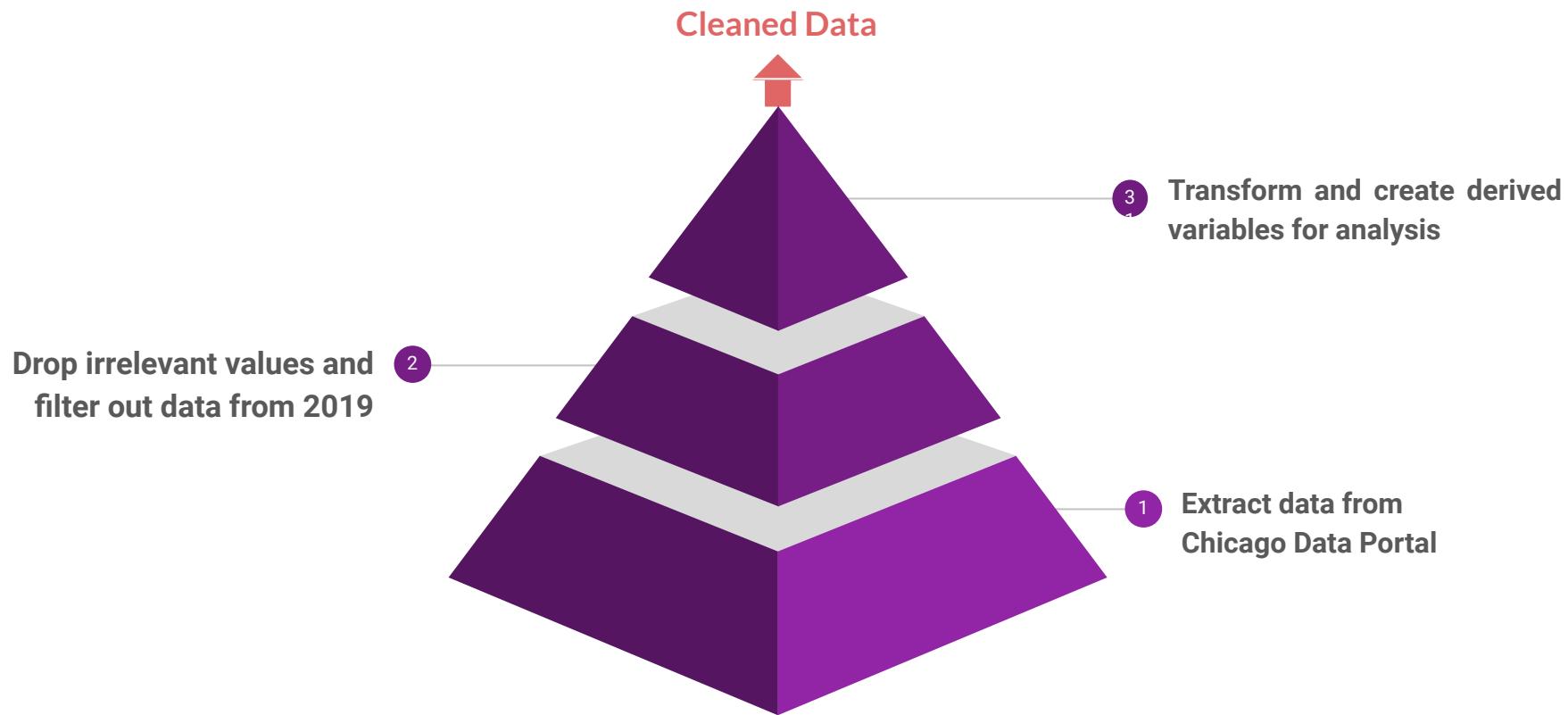
VEHICLE\_DEFECT

# Traffic Crash - Vehicle

- This dataset contains information about vehicles (or units as they are identified in crash reports) involved in a traffic crash.
- “Vehicle” information includes motor vehicle and non-motor vehicle modes of transportation, such as bicycles and pedestrians.
- Each mode of transportation involved in a crash is a “unit” and get one entry here.

# Data Preparation Steps: The ETL Process

---



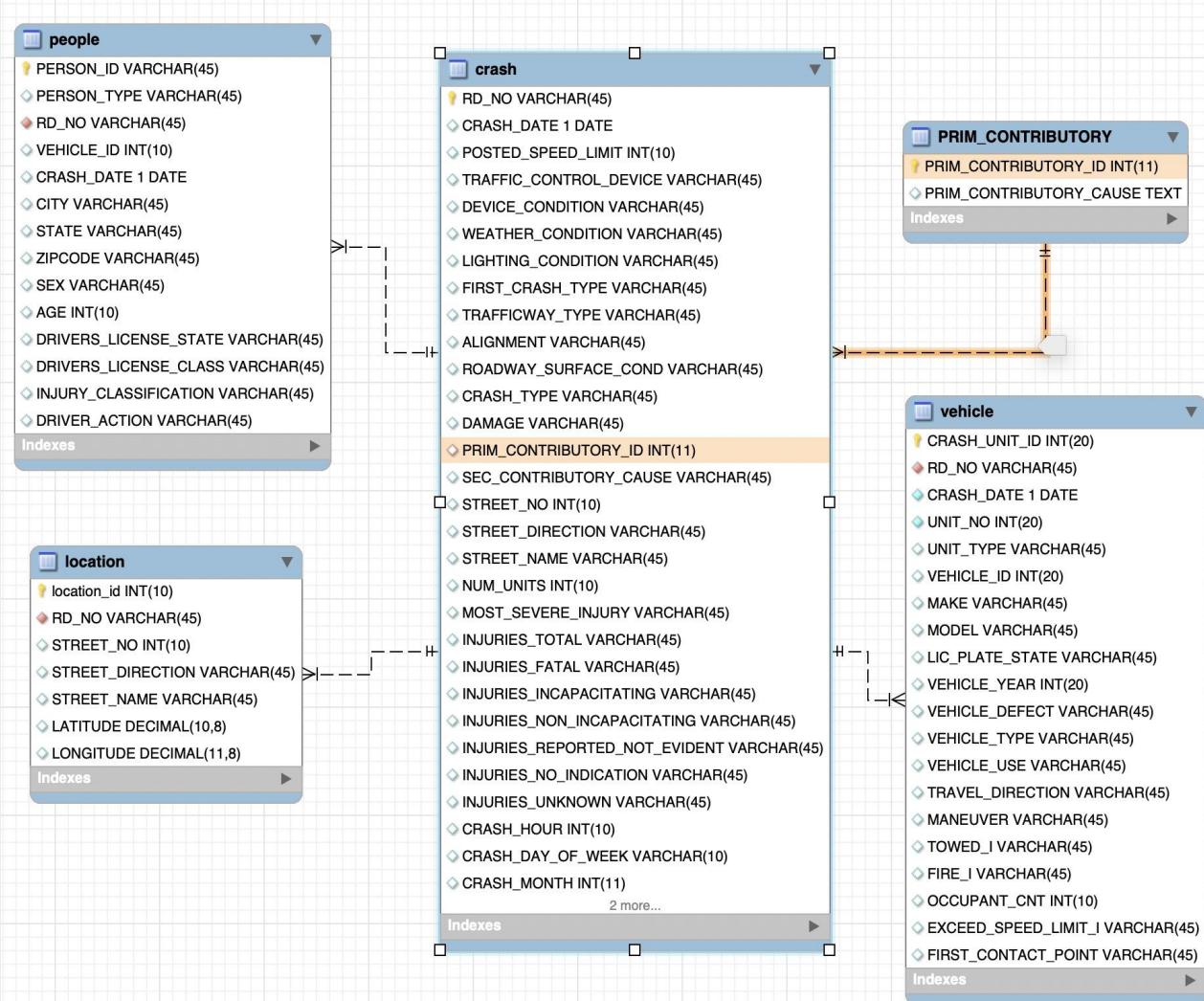
# Data Preparation Steps: Details

The purpose of

Crashes	CRASH_DATE	POSTED_SPEED	TRAFFIC_CONT
Filtered and selected data in 2019	crash<-crash[crash\$CRASH_MONTH<11, ]		
Split datetime into date and time to create hour,day,& month variables	crash[crash\$CRASH_DAY_OF_WEEK==1,]\$CRASH_DAY_OF_WEEK <- "Sun"		
Created zipcode variable from location	crash[crash\$CRASH_DAY_OF_WEEK==2,]\$CRASH_DAY_OF_WEEK <- "Mon"		
Recoade Prim_Contributor variable	crash[crash\$CRASH_DAY_OF_WEEK==3,]\$CRASH_DAY_OF_WEEK <- "Tue"		
Total of 98,169 records loaded into database	crash[crash\$CRASH_DAY_OF_WEEK==4,]\$CRASH_DAY_OF_WEEK <- "Wed"		
	crash[crash\$CRASH_DAY_OF_WEEK==5,]\$CRASH_DAY_OF_WEEK <- "Thu"		
	crash[crash\$CRASH_DAY_OF_WEEK==6,]\$CRASH_DAY_OF_WEEK <- "Fri"		
	crash[crash\$CRASH_DAY_OF_WEEK==7,]\$CRASH_DAY_OF_WEEK <- "Sat"		
	library(revgeo)		
	zipcode<-revgeo(crash\$LONGITUDE,crash\$LATITUDE,item='zip')		

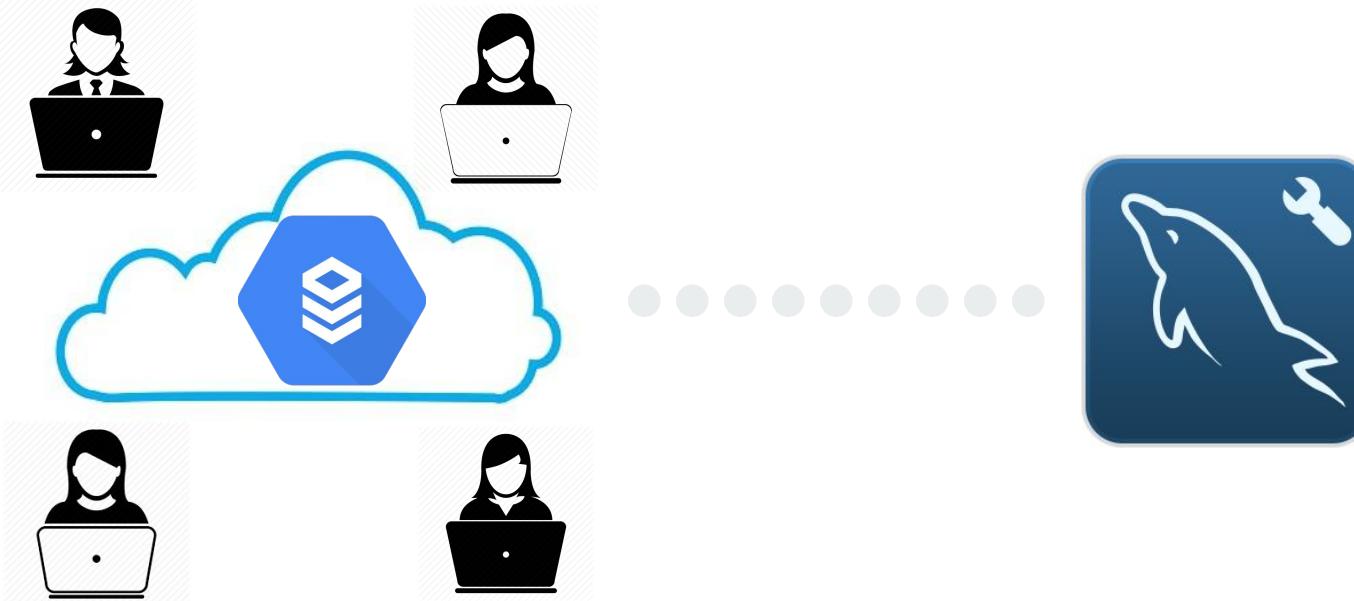
# DATABASE DESIGN

Enhanced Entity Relation (EER) Model



# Relational Database Implementation

---



# Relational Database Implementation : Create Database

```
17 • CREATE SCHEMA IF NOT EXISTS `traffic_crash` DEFAULT CHARACTER SET utf8;
18 • USE `traffic_crash` ;
19
20 -----
21 -- Table `traffic_crash`.`PRIM_CONTRIBUTORY`
22 -----
23 • ⏷ CREATE TABLE IF NOT EXISTS `traffic_crash`.`PRIM_CONTRIBUTORY` (
24     `PRIM_CONTRIBUTORY_ID` INT NOT NULL,
25     `PRIM_CONTRIBUTORY_CAUSE` TEXT,
26     PRIMARY KEY (`PRIM_CONTRIBUTORY_ID`)
27     ENGINE = InnoDB;
28
29 #CONSTRAINT `fk_PRIM_CONTRIBUTORY_crash`
30 # FOREIGN KEY (`PRIM_CONTRIBUTORY_ID`)
31 # REFERENCES `traffic_crash`.`PRIM_CONTRIBUTORY` (`PRIM_CONTRIBUTORY_ID`)
32 #ON DELETE NO ACTION
33 #ON UPDATE NO ACTION
34 -----
35 -- Table `traffic_crash`.`crash`
36 -----
37 • ⏷ CREATE TABLE IF NOT EXISTS `traffic_crash`.`crash` (
38     `RD_NO` VARCHAR(45) NOT NULL,
39     `CRASH_DATE` date,
40     `POSTED_SPEED_LIMIT` INT(10) NULL NULL,
41     `TRAFFIC_CONTROL_DEVICE` VARCHAR(45),
42     `DEVICE_CONDITION` VARCHAR(45),
43     `WEATHER_CONDITION` VARCHAR(45),
44     `LIGHTING_CONDITION` VARCHAR(45),
```



Traffic\_crash



# Relational Database Implementation: Load Data

The screenshot shows the Google Cloud Platform Storage interface. The left sidebar has a 'Storage' section with 'Browser' selected, followed by 'Transfer', 'Transfer Appliance', and 'Settings'. The main area shows 'Bucket details' for 'crash\_depa'. The top navigation bar includes 'My First Project' and search and refresh buttons. Below the bucket name, there are tabs for 'Objects', 'Overview', 'Permissions', and 'Bucket Lock', with 'Objects' being the active tab. A row of buttons provides actions: 'Upload files', 'Upload folder', 'Create folder', 'Manage holds', and 'Delete'. A search bar labeled 'Filter by prefix...' is present. The 'Buckets / crash\_depa' section lists four CSV files:

<input type="checkbox"/>	Name	Size	Type	Storage class	Last modified
<input type="checkbox"/>	TrafficCrashesVehicles2.csv	29.13 MB	text/csv	Standard	11/8/19, 3:6
<input type="checkbox"/>	crash2.csv	33.46 MB	text/csv	Standard	11/8/19, 2:6
<input type="checkbox"/>	crash3-csv.csv	31.36 MB	text/csv	Standard	11/8/19, 4:6
<input type="checkbox"/>	crash4.csv	31.31 MB	text/csv	Standard	11/8/19, 1: UTC-6

# Relational Database Implementation: Load Data

## Import data from Cloud Storage

### Source

#### Choose the file you'd like to import data from

Browse for a file, or enter the path for one (bucket/folder/file). Make sure you have read access first. [Learn more](#)

 crash\_depa/crash4.csv[Browse](#)

#### Indicate the format of the file you're importing

 SQL

A plain text file with a sequence of SQL commands, like the output of mysqldump

 CSV

If your Cloud Storage file is a CSV file, select CSV. The CSV file should be a plain text file with one line per row and comma-separated fields.

### Destination

Choose the database and table in your Cloud SQL instance that you'd like to import your file into

#### Database

 traffic\_crash

#### Table

 crash[Import](#)

When you import, a Cloud SQL service account will be granted read access to your Cloud Storage file and the bucket that contains it. This will be reflected in your permissions.

RD_NO	CRASH_DATE_1	POSTED_SPEED_LIMIT	TRAFFIC_CONTROL_DEVICE	DEVICE_CONDITION	WEATHER_CONDITION	LIGHTING_CONDITION
BF358140	2019-07-21	20	NO CONTROLS	NO CONTROLS	CLOUDY/OVERCAST	UNKNOWN
CJ114279	2019-01-12	25	STOP SIGN/FLASHER	NO CONTROLS	SNOW	DAYLIGHT
CJ114864	2019-01-12	30	STOP SIGN/FLASHER	FUNCTIONING PROPERLY	SNOW	DARKNESS
CJ132641	2019-01-28	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
CJ132951	2019-01-28	30	NO CONTROLS	NO CONTROLS	SNOW	DAYLIGHT
CJ162000	2019-02-21	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
CJ176557	2019-03-07	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
CJ391262	2019-08-10	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
CJ396168	2019-08-17	30	NO CONTROLS	NO CONTROLS	RAIN	DAYLIGHT
DJ429730	2019-09-12	30	TRAFFIC SIGNAL	FUNCTIONING PROPERLY	CLEAR	DARKNESS, LIGHTED ROAD
HC180501	2019-03-10	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
JA112132	2019-01-10	30	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD
JA144527	2019-02-08	30	TRAFFIC SIGNAL	FUNCTIONING PROPERLY	CLEAR	DAYLIGHT
JA150613	2019-02-13	30	NO CONTROLS	NO CONTROLS	SNOW	DAYLIGHT
JA236480	2019-04-24	30	TRAFFIC SIGNAL	FUNCTIONING PROPERLY	CLEAR	DAYLIGHT
JA366308	2019-07-27	30	TRAFFIC SIGNAL	FUNCTIONING PROPERLY	CLEAR	DARKNESS, LIGHTED ROAD
JB101727	2019-01-02	35	NO CONTROLS	NO CONTROLS	CLOUDY/OVERCAST	DAYLIGHT
JB103018	2019-01-03	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
JB103093	2019-01-03	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
JB103281	2019-01-03	35	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD
JB104200	2019-01-03	30	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS
JB104437	2019-01-04	20	TRAFFIC SIGNAL	FUNCTIONING PROPERLY	CLEAR	DAYLIGHT
JB105170	2019-01-04	30	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD
JB105531	2019-01-05	35	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
JB106020	2019-01-01	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
JB106860	2019-01-06	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
JB109139	2019-01-08	20	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT

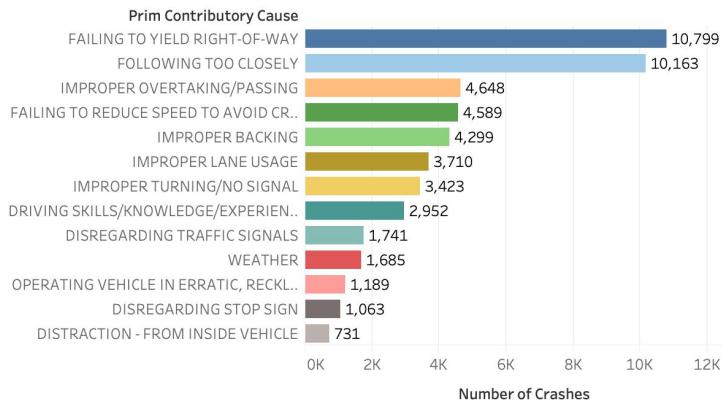
# Analysis & Visualizations

---

# Crashes: Contributor & Damage Level

---

Top 13 Primary Contributors



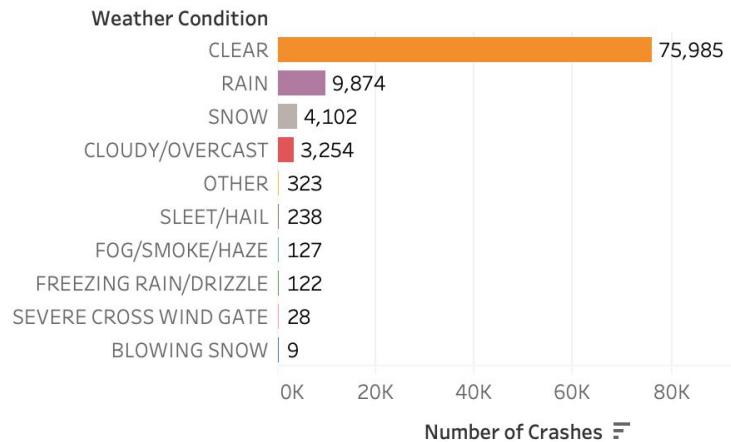
Damage Level



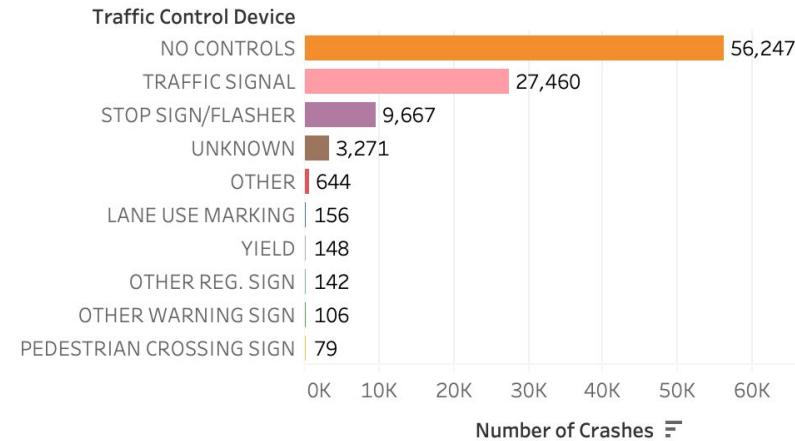
- Failing to yield right-of-way is the top 1 contributor to all traffic crashes.
- Most accidents are not caused by external forces but rather man-made errors associated with bad driving behavior.
- Most crashes result in damage exceeding \$1,500.

# Crashes and Different Conditions

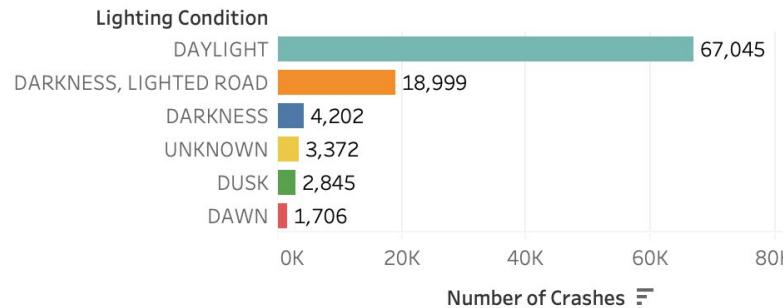
Weather Condition and Crashes



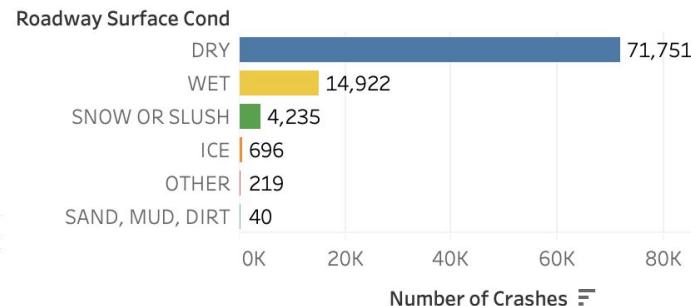
Traffic Control Devices and Crashes



Lighting Condition and Crashes



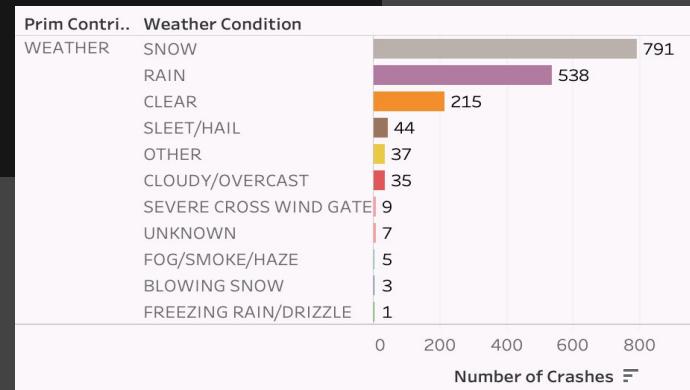
Roadway Surface Condition and Crashes



# Crash: SQL Query

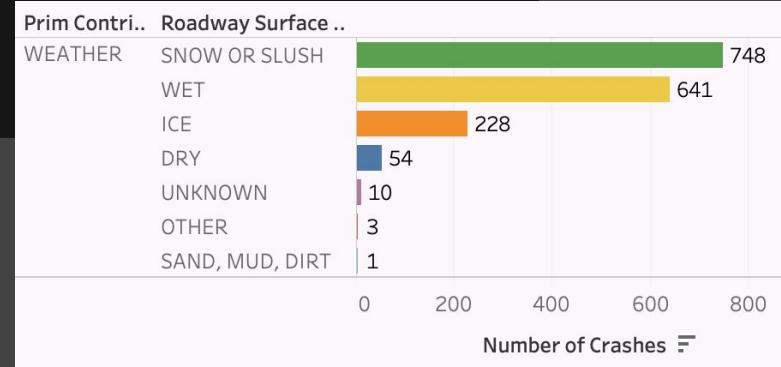
---

```
# Query weather and crash count from crashes whose primary contributor is weather
SELECT
    c.weather_condition, COUNT(*) AS crashCount
FROM
    cause_count cc
        INNER JOIN
    crash c ON cc.prim_contributory_id = c.prim_contributory_id
WHERE
    cc.prim_contributory_cause = 'WEATHER'
GROUP BY weather_condition
ORDER BY crashCount DESC;
```



# Crash: SQL Query

```
# Query the road surface condition and crash count when the primary contributor is weather
SELECT
    c.roadway_surface_cond,
    count(*)
FROM
    cause_count cc
    INNER JOIN
    crash c ON cc.prim_contributory_id = c.prim_contributory_id
WHERE
    cc.prim_contributory_cause = 'WEATHER'
GROUP BY c.roadway_surface_cond
ORDER BY count(*) DESC;
```



# Crash: SQL Query

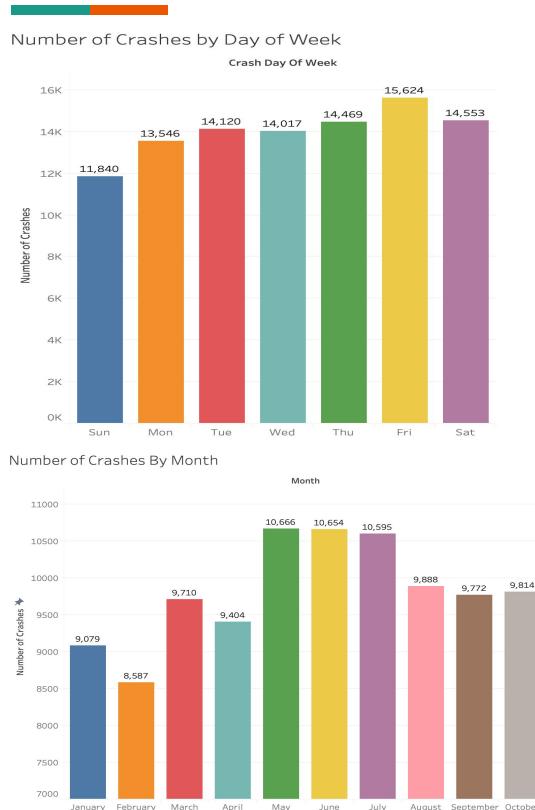
---

```
# Query the lighting condition and crash count when the primary contributor is weather
SELECT
    c.lighting_condition,
    COUNT(*) as crash_count
FROM
    cause_count cc
        INNER JOIN
    crash c ON cc.prim_contributory_id = c.prim_contributory_id
WHERE
    cc.prim_contributory_cause = 'WEATHER'
GROUP BY c.lighting_condition
ORDER BY crash_count DESC;
```

Lighting vs Crashes

lighting_condition	crash_count
DAYLIGHT	962
DARKNESS, LIGHTED ROAD	488
DARKNESS	111
DUSK	63
DAWN	44
UNKNOWN	17

# Crashes and Timing

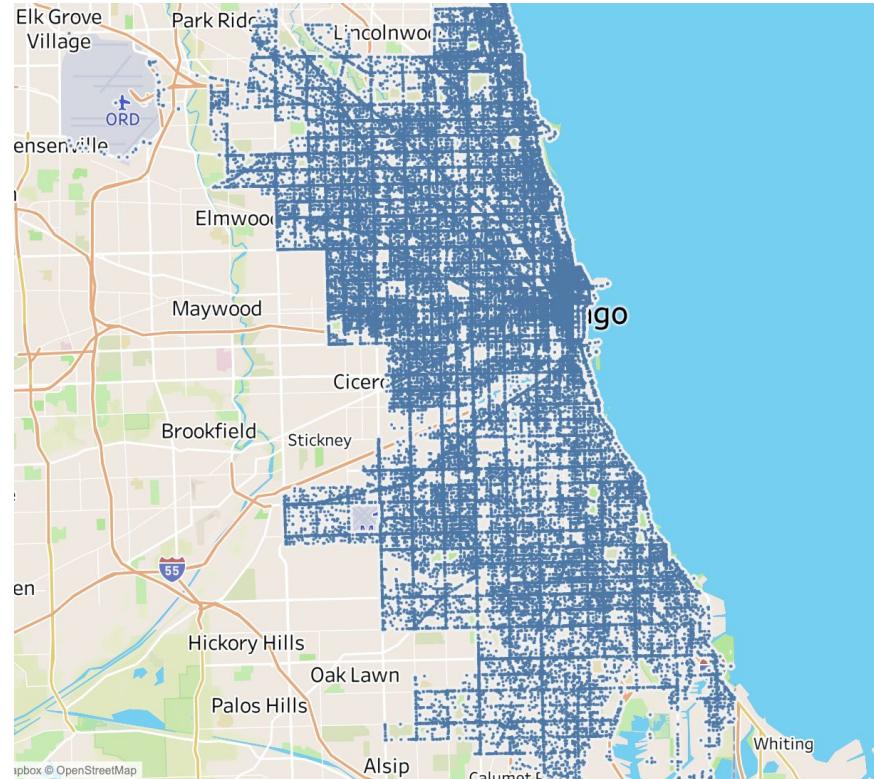


- Most accidents take place in the afternoon peak hours (3-5pm).
- Most accidents take place on the weekend. (Thursday-Saturday)
- Most accidents happen in the summer months. (Note : our data only covers up to end of October. )



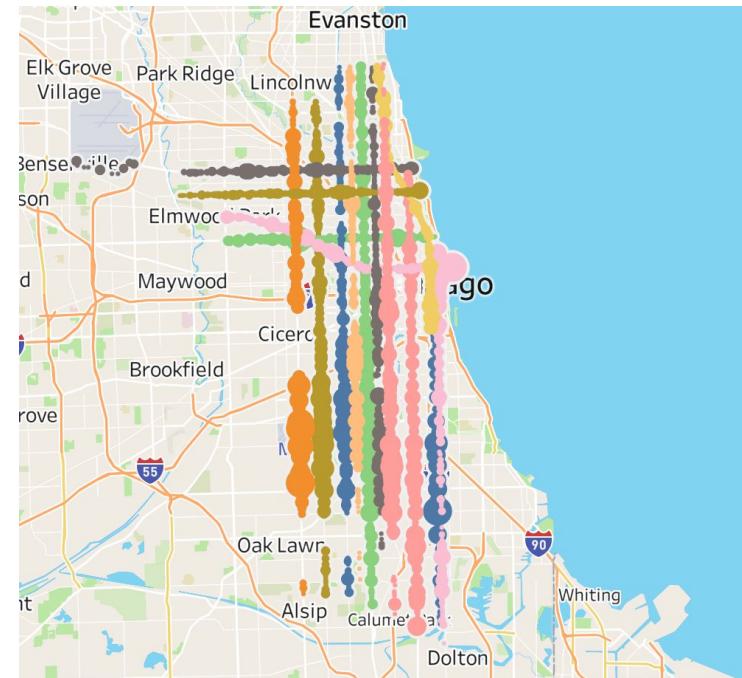
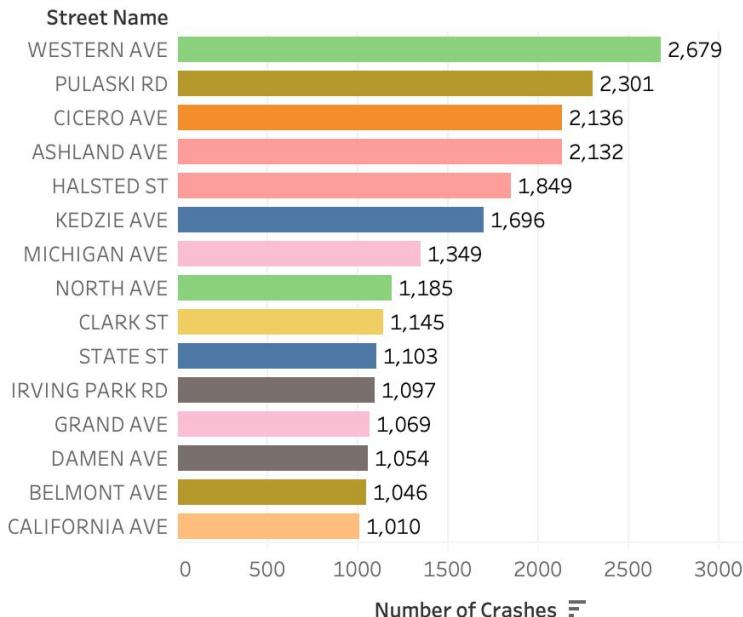
# Crash Location Visualization

Most traffic crashes happen in downtown Chicago area.



# Crashes and the Streets

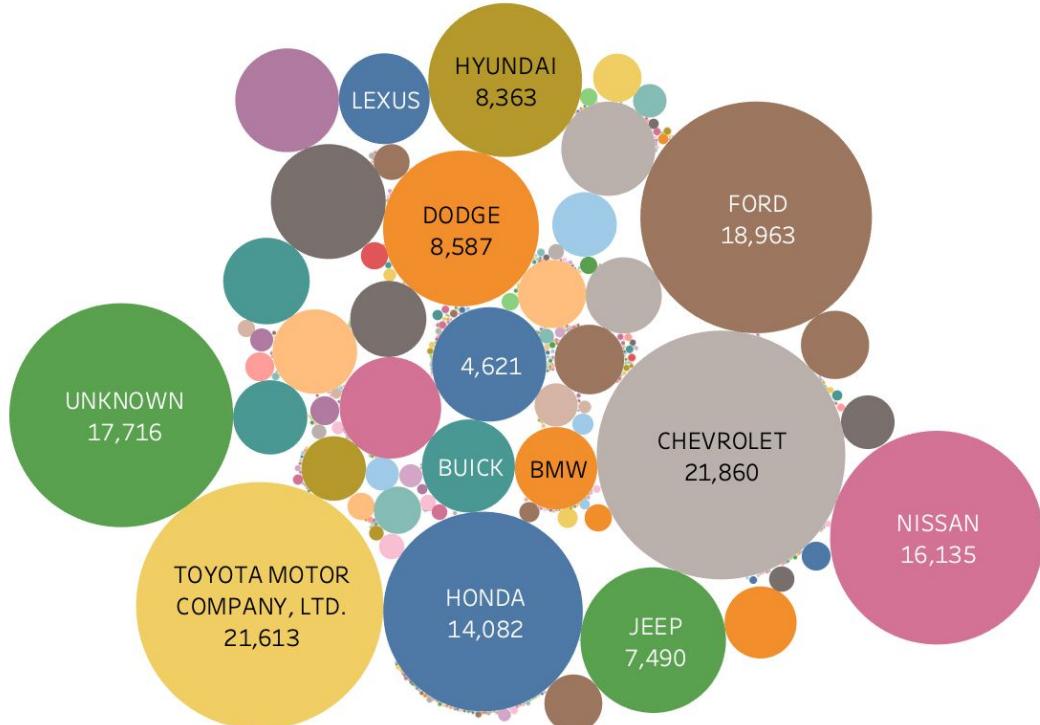
## Top 15 Streets with Most Number of Crashes



# Vehicle: Vehicle Make

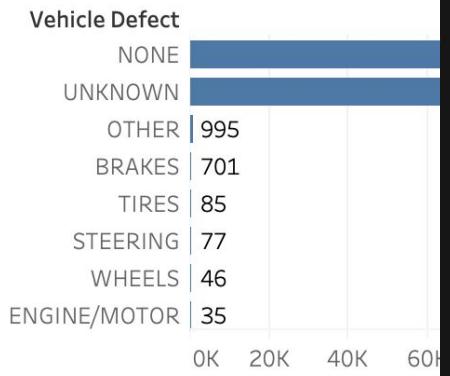
---

- Top 5 vehicle makes (>10000) involved in accidents in Chicago are:  
**CHEVROLET**  
**TOYOTA**  
**FORD**  
**NISSAN**  
**HONDA**
- But this is not subjective as these are the most popular car makes in the US.
- We found no relationship between makes and number of crashes.



# Vehicle: Vehicle Defect

## Vehicle Defect



```
# Query vehicle defect and crash count from vehicle defect table
SELECT
    v.vehicle_defect,
    COUNT(*) AS crashCount
FROM
    vehicle v
        INNER JOIN
    crash c ON v.RD_NO = c.RD_NO
        INNER JOIN
    cause_count cc ON cc.prim_contributory_cause = 'EQUIPMENT'
WHERE
    cc.prim_contributory_cause = 'EQUIPMENT'
GROUP BY vehicle_defect
ORDER BY crashCount DESC;
```

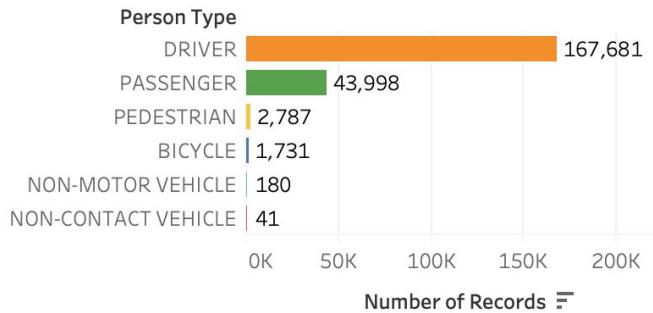
## Vehicle Defect vs Crashes

vehicle_defect	crashCount
NONE	527
UNKNOWN	282
BRAKES	236
OTHER	62
STEERING	40
TIRES	37
WHEELS	21

- Most vehicles involved in the car accident have no defect or defect is unknown.
- Brake failure is the most common issue in known vehicle defect.

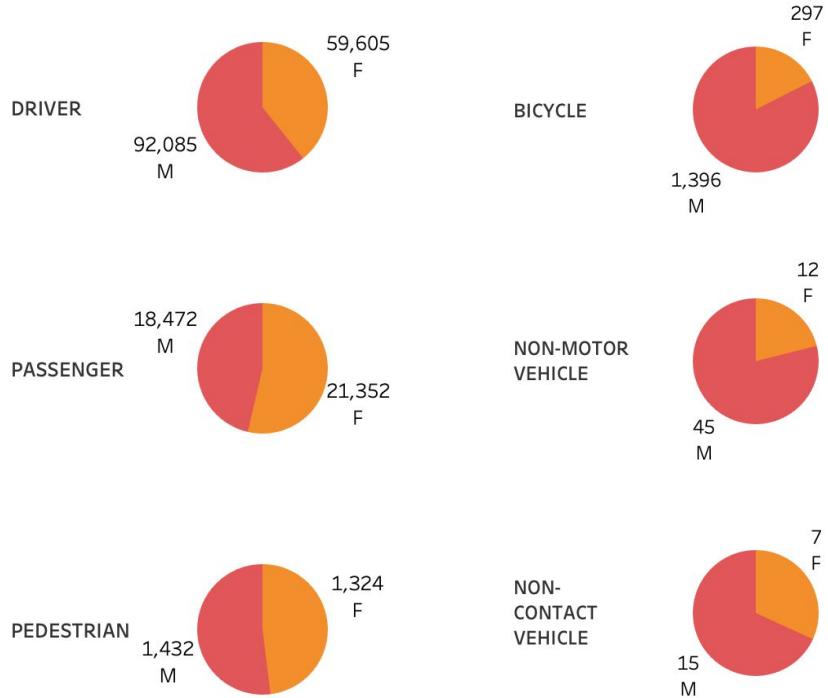
# People: Person Type

## Person Type



- Most people involved in traffic crashes are drivers.
- There are more male drivers than female drivers.

Person Type



# People: SQL Query

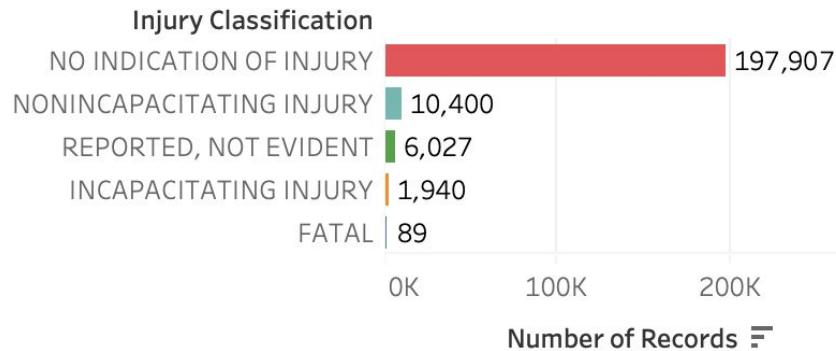
```
# Query sex of drivers and the number of accidents caused by manmade errors
● SELECT
    p.SEX,
    COUNT(c.RD_NO) AS manmade_crash_count
  FROM
    cause_count cc
    INNER JOIN
    crash c ON cc.prim_contributory_id = c.prim_contributory_id
    INNER JOIN
    people p ON c.RD_NO = p.RD_NO
 WHERE
    prim_contributory_cause IN ('FAILING TO YIELD RIGHT-OF-WAY',
    AND person_type = "DRIVER"
 GROUP BY p.SEX
 ORDER BY manmade_crash_count DESC
 LIMIT 2;
```

Sex of Drivers	
SEX	manmade_crash_count
M	53469
F	34604

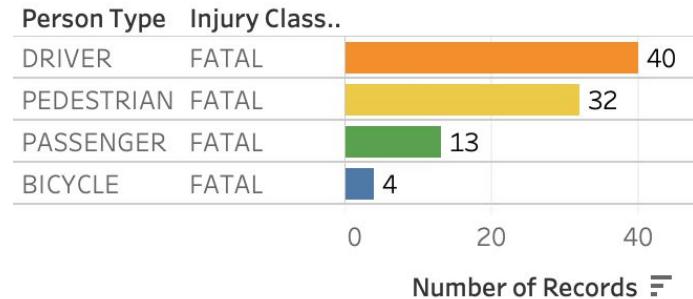
# People: Injury Classification

---

## Injury Classification



## Injury:PersonType



- Most crashes cause no injury.
- Among the fatal crashes, drivers and pedestrians are most at risk of death.

---

# Recommendations

For commuters

- Practice good driving habits, especially for male drivers!
- Stay cautious while driving - accidents tend to happen when you let loose.
- Review traffic rules and regulations from time to time.

For government authorities

- Set up traffic control devices and warning signs at high-risk locations
- Allot more personnel during the afternoon peak hours and on Fridays and Saturdays

---

# Lessons Learned

- Choosing the right tools can save a bunch of time!  
(e.g. OpenRefine vs R, load data by local SQL vs cloud SQL)
- Always have problems in mind before data preparation. Otherwise you might need to reload data into your database or create new variables again and again.
- Data preparation can be more time-consuming than analysis and modeling. Clean and tidy datasets can ease the analysis process.
- Let the data tell its story. Our preconception might be completely overturned.

# Thank you!

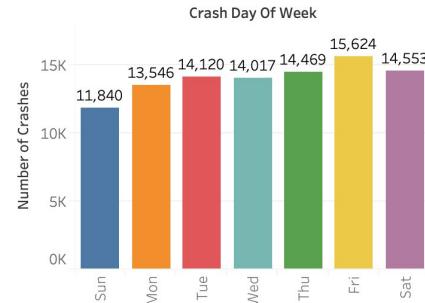
Q & A

---

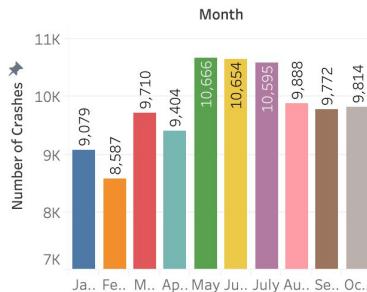
# Appendix 1: Timing Dashboard

---

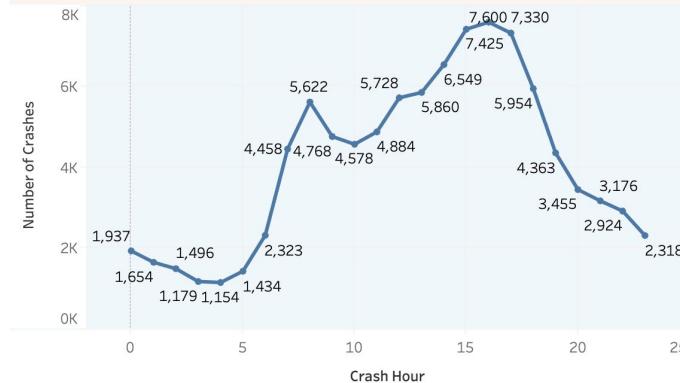
Number of Crashes by Day of Week



Number of Crashes By Month



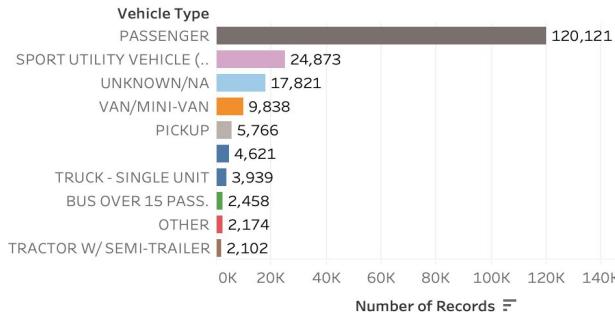
Number of Crashes by Hour of Day



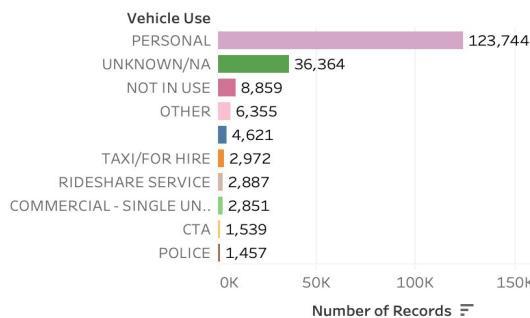
# Appendix 2: Vehicle Dashboard



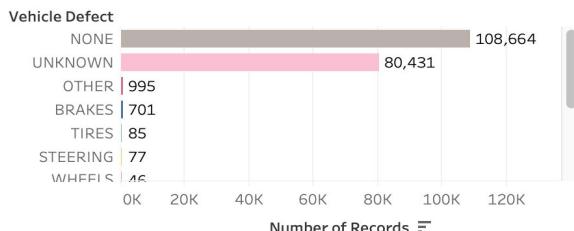
## Type



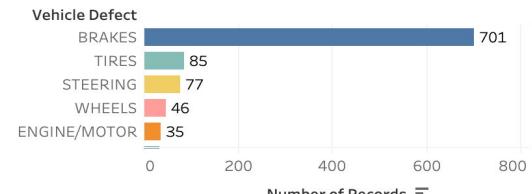
## Use



## Vehicle Defect

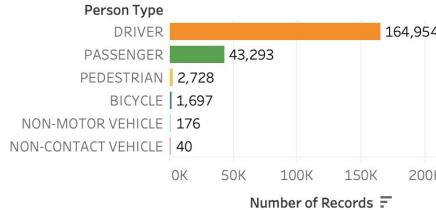


## Number of Crashes for Vehicle Defect Known



# Appendix 3: People Dashboard

## Person Type



## Injury Classification



## Injury:PersonType



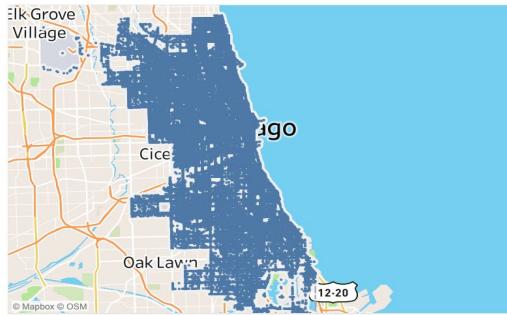
## PersonType:Sex



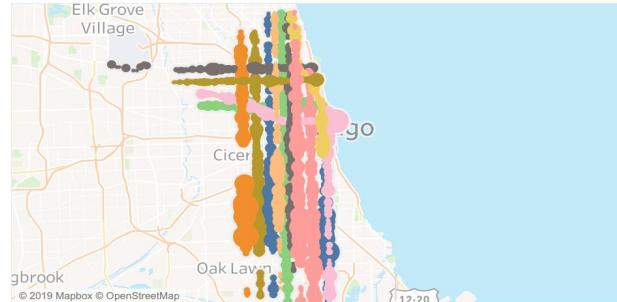
# Appendix 4: Location Dashboard

---

Map



Street and Number of Crashes



Top 10 Street with Most Number of Crashes

