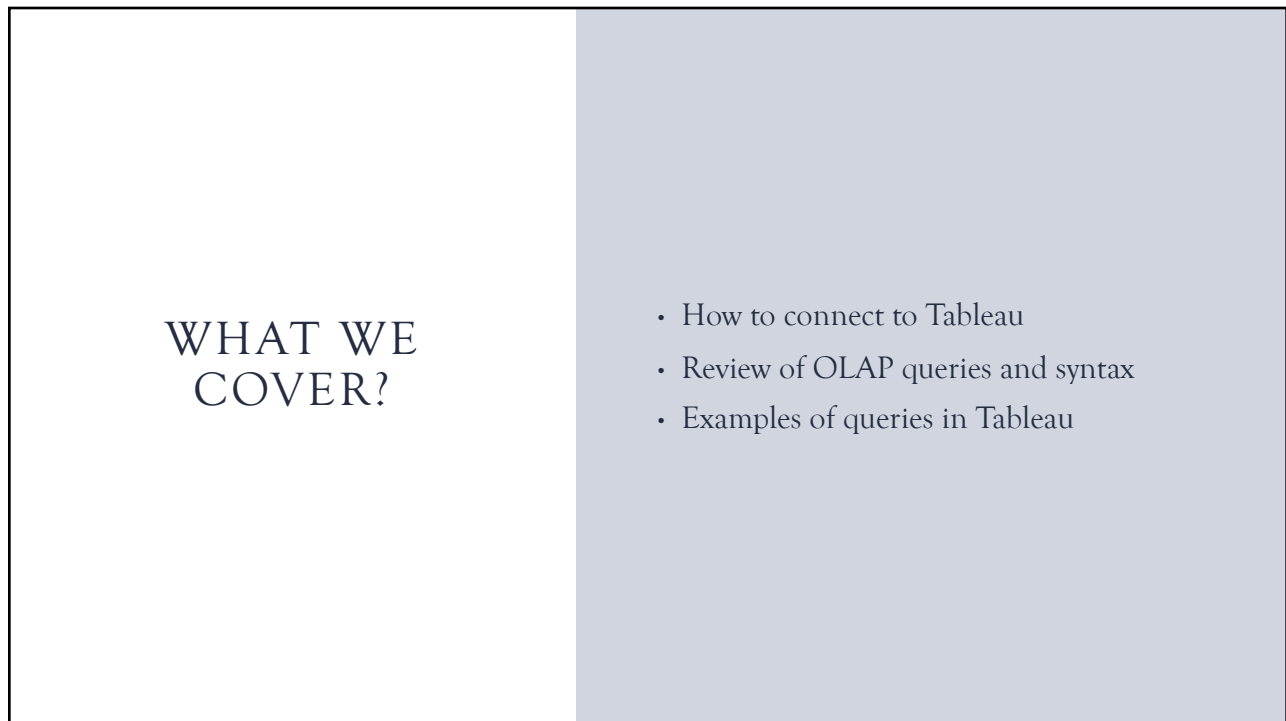




1



2



## HOW TO CONNECT TO TABLEAU

- [Download the latest version of Tableau Desktop and Tableau Prep Builder here](#)
- Click on the link above and select “Download Tableau Desktop” On the form, enter your school email address for Business E-mail and enter University of Ottawa as organization.
- Activate with your product key: TC27-F858-A340-FBE5-76A2

3

## FIRST, CONNECTING PG-SQL TO TABLEAU

- Select the connect to server.
- Choose Postgres SQL
- Add the server, port, group database, username and password
- Check the require SSL
- Now you should be able to connect!

PostgreSQL

General Initial SQL

Server  
www.eecs.uottawa.ca

Port  
15432

Database  
group\_

Authentication  
Username and Password

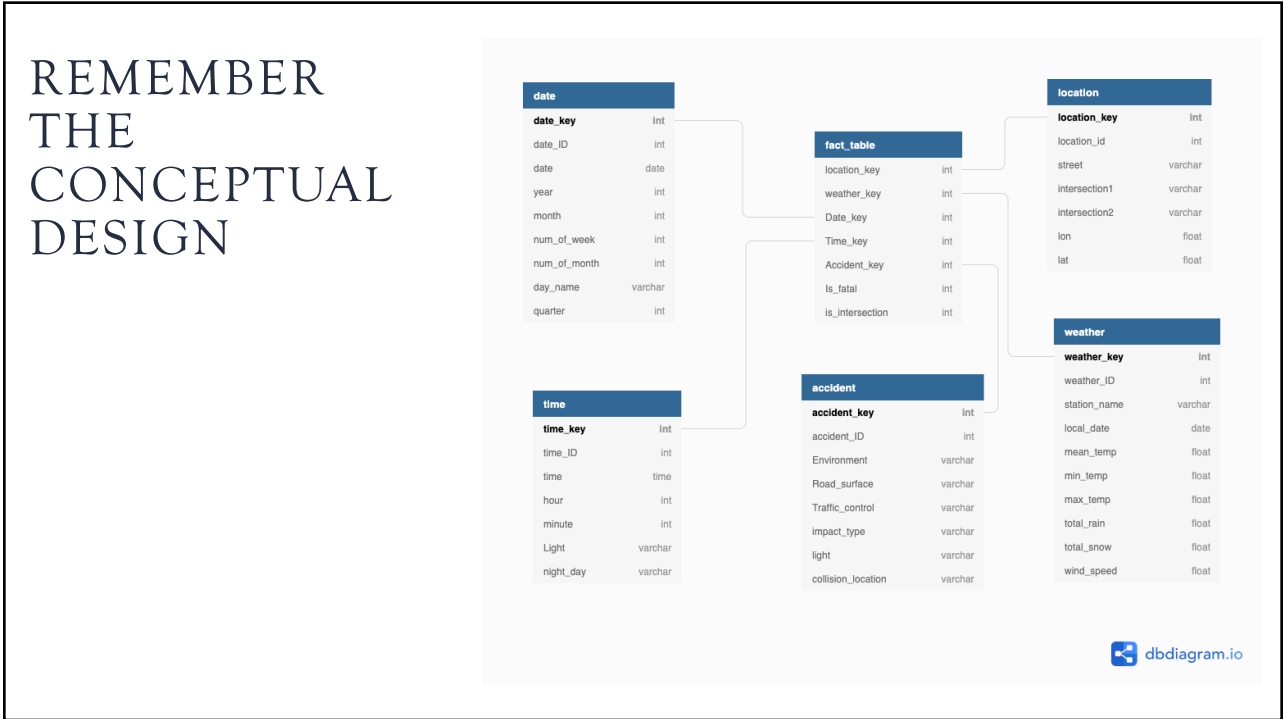
Username

Password  
Optional

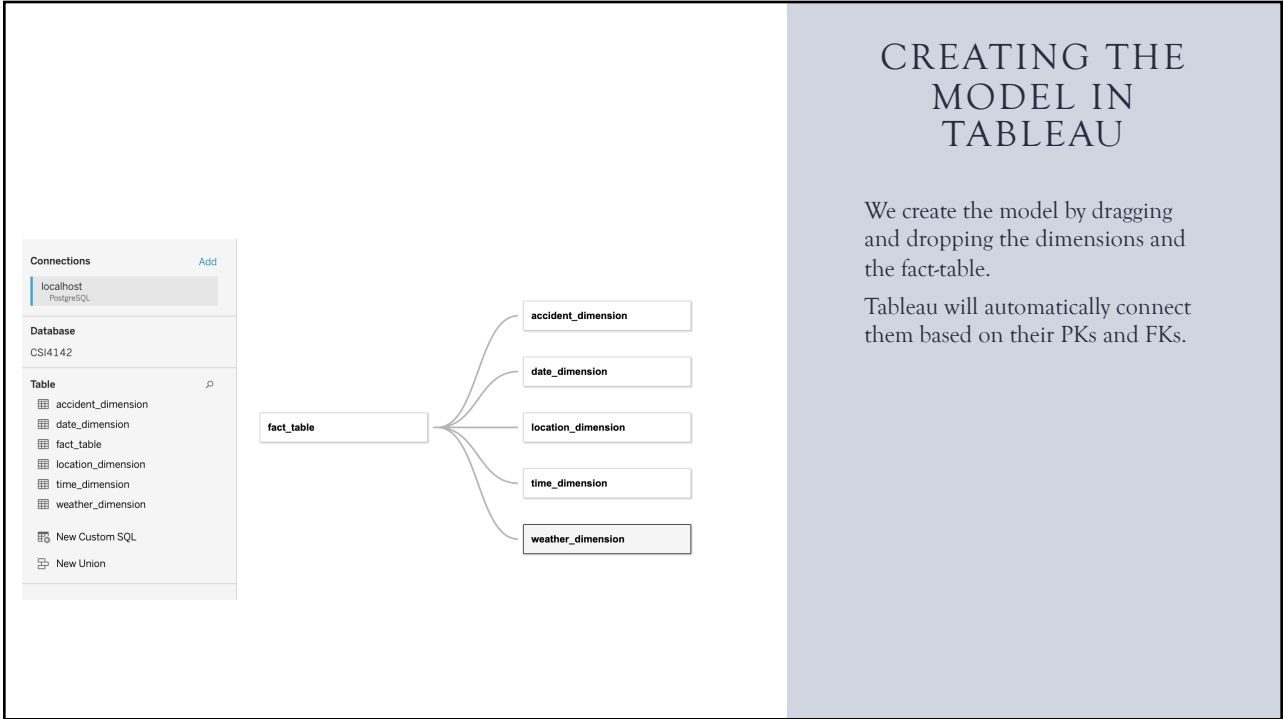
☒ Require SSL

Sign In

4



5



6

## WHAT IS NEEDED FOR THE PROJECT



OLAP queries in SQL [60]



BI dashboard and data  
visualization (Tableau) [40]

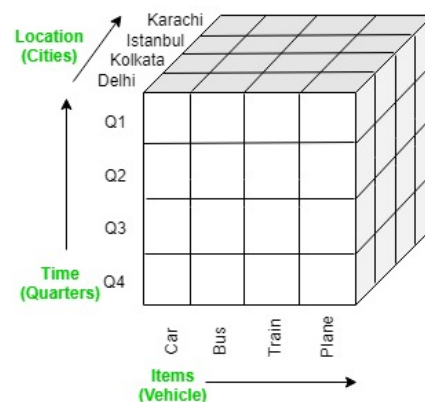
7

## CUBE

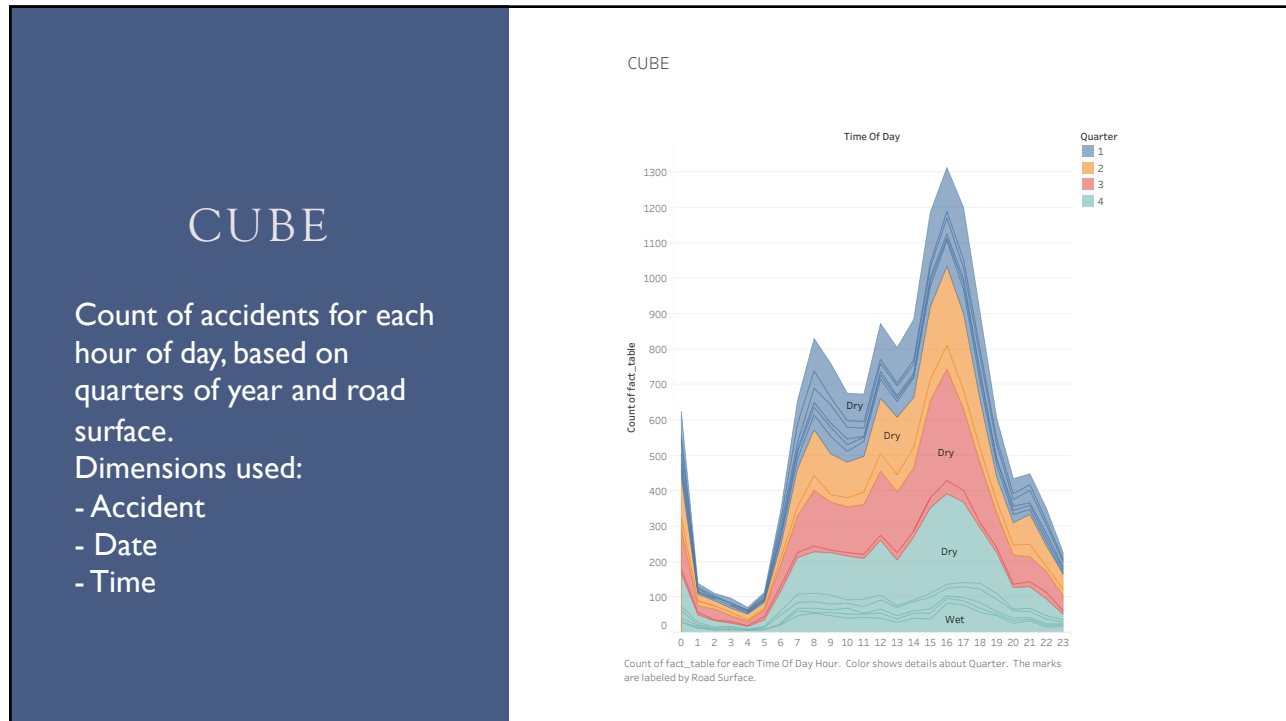
CUBE: A multi-dimensional array of data.

In the example, we consider all locations,  
all items in the four quarters of the year.

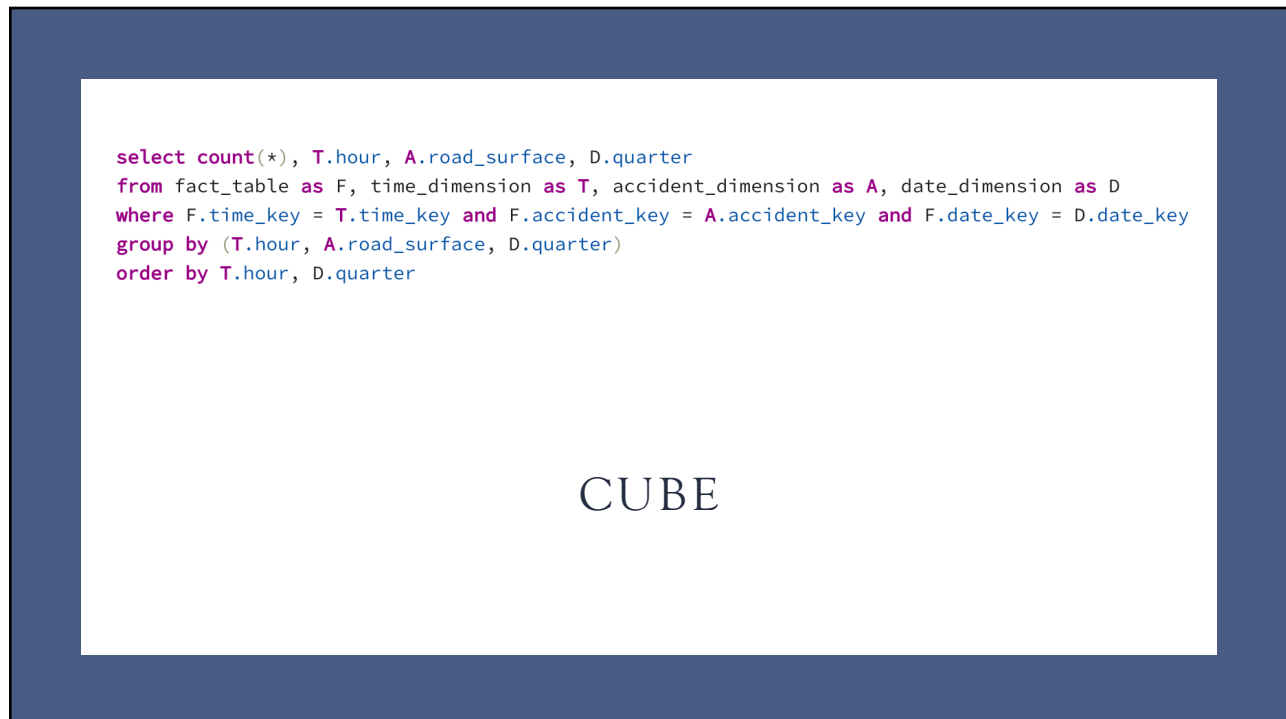
\*Picture from <https://www.geeksforgeeks.org/olap-operations-in-dbms/>



8



9



10

## HOW TO DO IT IN TABLEAU

- For columns, we only add the hour of day.
- For rows we add the number of accidents.
- In the marks, we set the quarter to color, and the road surface to labels (in order to better represent them in the figure).

11

## SLICE

- Definition of slice: We filter based on one dimension.
- In the example here, we set the time dimension to be Q1.

\*Picture from [https://www.tutorialspoint.com/dwh/dwh\\_olap.htm](https://www.tutorialspoint.com/dwh/dwh_olap.htm)

The diagram shows a 3D cube representing data across three dimensions: Locations (cities), Time (Quarter), and item(types). The cube is sliced for Time = Q1, resulting in a 2D table showing data for Q1 only.

Locations (cities)	Mobile	Modem	Phone	Security
Chicago	605	825	14	400
New York				
Toronto				
Vancouver				

12



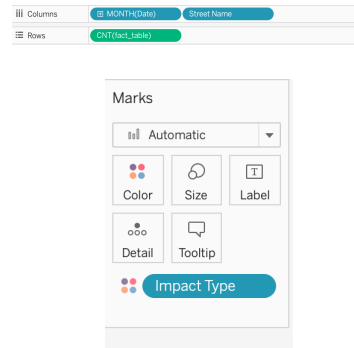
13



14

## HOW TO DO IT IN TABLEAU

1. For columns, we add the months and street.
2. For rows we add the number of accidents.
3. Since we are slicing based on street, we filter street to Bank Street.
4. In the marks, we set the impact type to Color to present all the impact types.

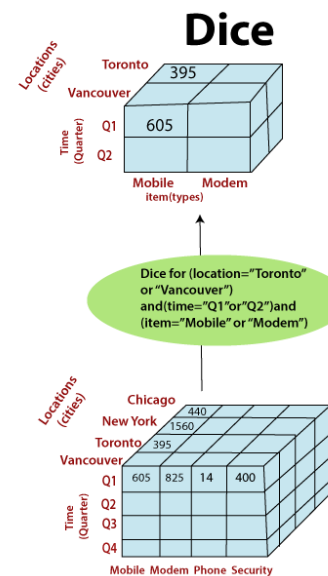


15

## DICE

- Definition of Dice: We create a Subcube.
- In the example, we are filtering based on all dimensions.

\*Picture from <https://www.javatpoint.com/olap-operations>



16

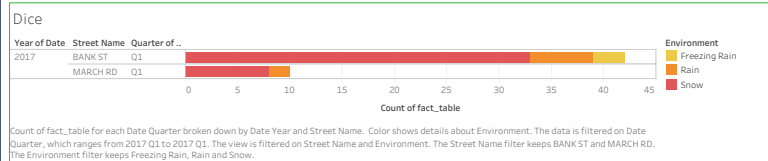


## DICE

*Count of accidents happening in Freezing rain, Rain and Snowy environments on Bank and March in Q1 2017.*

Dimensions used:

- Accident (Snow and Rain)
- Location (Bank and March Streets)
- Date (Q1 2017)



17

```
select count(*), L.street_name, A.environment, D.quarter
from fact_table as F, location_dimension as L, accident_dimension as A, date_dimension as D
where F.location_key = L.location_key and F.accident_key = A.accident_key and F.date_key = D.date_key
and L.street_name in ('BANK ST','MARCH RD') and D.quarter = 1 and A.environment in ('Rain','Snow','Freezing Rain')
group by (L.street_name, A.environment, D.quarter)
```

## DICE

18

## HOW TO DO IT IN TABLEAU

1. For columns, we add count of fact-table.
2. For rows we add the quarter, street names.
3. In the marks, we set the Environment to Color to present all the environment types.
4. We filter all dimensions. For street, we filter based on March and Bank Streets. For Date, we filter on Q1. Finally, for Environment, we filter based on Rain and Snow.

Tableau interface showing Marks and Filters shelves. The Marks shelf has 'Environment' as the color. The Filters shelf has 'Street Name', 'QUARTER(Date)', and 'Environment'. Below, the Columns shelf shows 'CNT(fact\_table)' and the Rows shelf shows 'YEAR(Date)', 'Street Name', and 'QUARTER(Date)'.

19

## DRILL DOWN

- Drill down: Opposite of roll-up, i.e. when we zoom in to more details.
- Example: Drilling down from Quarters to Months.

\*Picture from <https://www.javatpoint.com/olap-operations>

Diagram illustrating Drill Down. It shows a 3D cube for 'Time (Quarter)' with dimensions 'Location (City)' and 'Fact (Mobile, Modem, Phone, Security, ItemTypes)'. The cube is divided into quarters (Q1, Q2, Q3, Q4). An arrow labeled 'Drilldown on time from quarters to month' points to a more detailed 3D cube showing months (Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec) for the same dimensions.

20

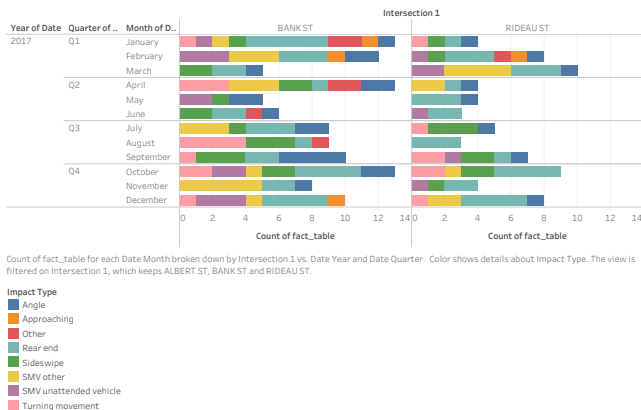
## DRILLING DOWN

*Count of accidents happened in Bank and Rideau intersections grouped by months of the year..*

*Dimensions used:*

- Accident (Impact type)
- Location (Intersection of Bank and Rideau)
- Date (Drill down based on Month)

Drill down



21

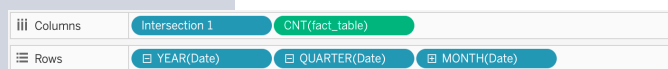
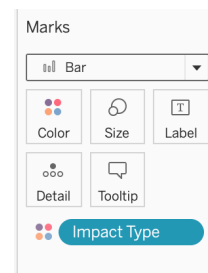
```
select count(*), L.intersection_1, A.impact_type, D.quarter, D.month_name, D.year
from fact_table as F, location_dimension as L, accident_dimension as A, date_dimension as D
where F.location_key = L.location_key and F.accident_key = A.accident_key and F.date_key = D.date_key
and L.intersection_1 in ('BANK ST','RIDEAU ST')
group by (L.intersection_1, A.impact_type, D.quarter, D.month_name, D.year)
order by L.intersection_1, D.quarter, D.month_name, D.year
```

## DRILLING DOWN

22

## HOW TO DO IT IN TABLEAU

1. For columns, we add year, quarter and month. (Remember to keep the order for showing the concept hierarchy.)
2. For rows we add intersection and accident count.
3. In the marks, we set Impact to Color to present all the impact types.
4. We need to filter the intersection of Bank and Rideau streets.



23

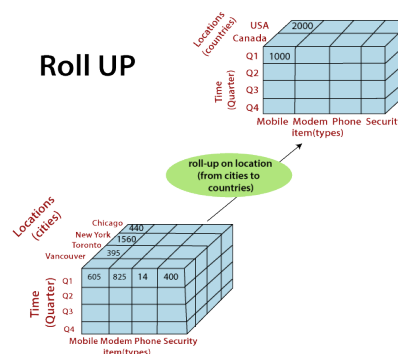
## ROLL-UP

Roll-up: Aggregating the data cube by climbing up the hierarchies.

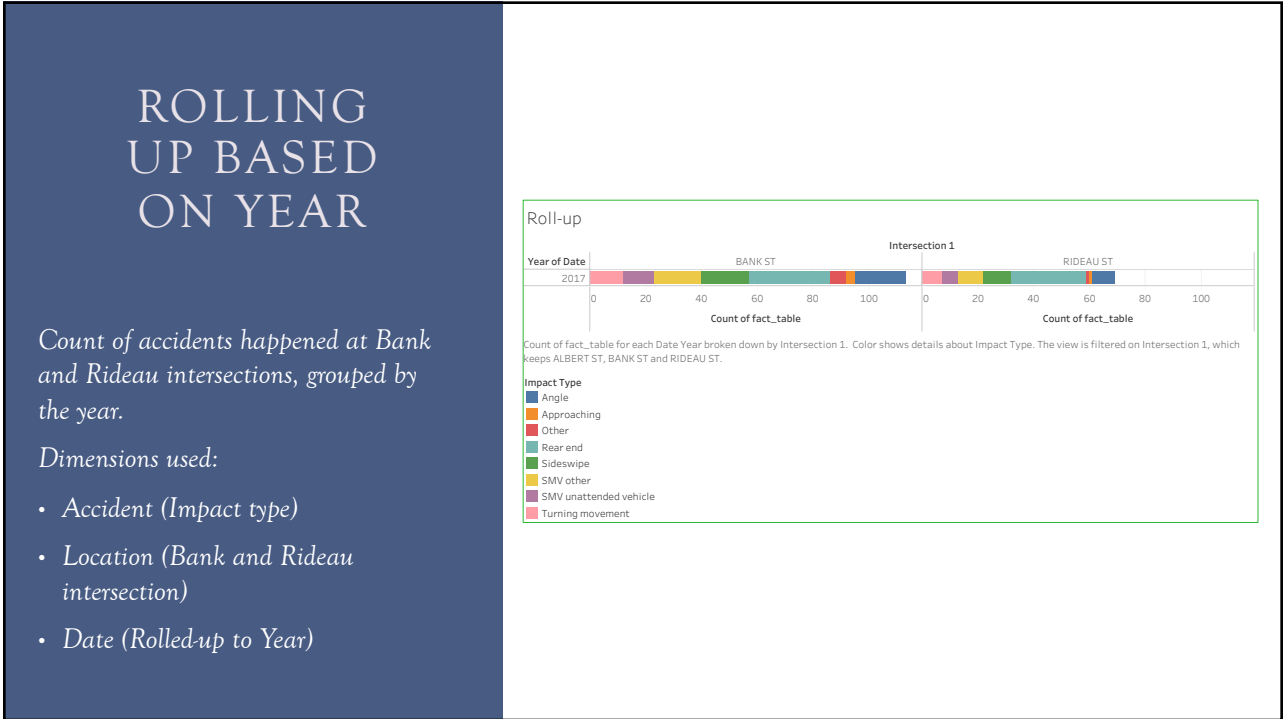
Examples:

1. street, city, country
2. day, month, year

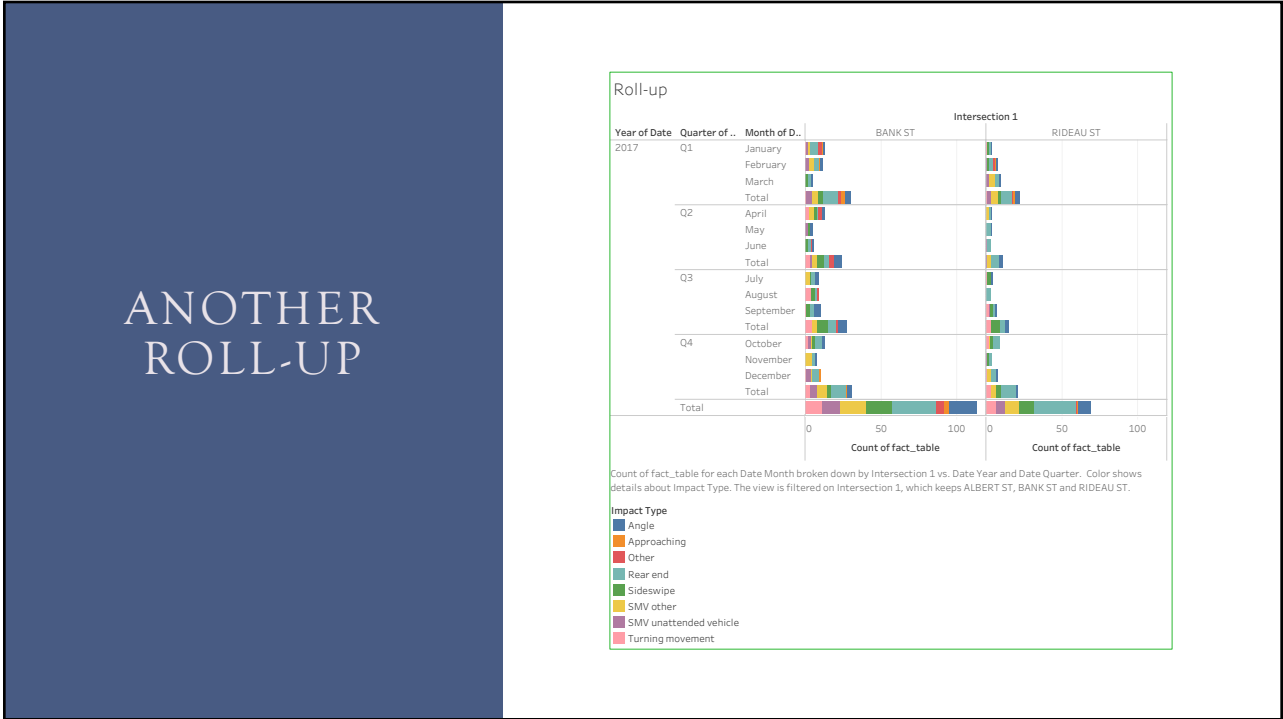
\*Picture from <https://www.javatpoint.com/olap-operations>



24



25



26

```
select count(*), D.year, D.quarter, D.month_name, A.impact_type, L.intersection_1
from fact_table as F, date_dimension as D, location_dimension as L, accident_dimension as A
where F.date_key = D.date_key and F.location_key = L.location_key and F.accident_key = A.accident_key
and L.intersection_1 in ('BANK ST','RIDEAU ST')
group by A.impact_type, L.intersection_1, rollup (D.year, D.quarter, D.month_name)
order by D.year, D.quarter, D.month_name
```

ROLL-UP

27

ICEBERGS

Top 10 streets with the most accidents in 2017

Dimensions used:

- Location
- Date

ICE berg

Street Name	Count of fact_table
HIGHWAY 417	1,484
BANK ST	1,484
CARLING AVE	1,484
INNES RD	1,484
BASELINE RD	1,484
GREENBANK RD	1,484
PRINCE OF WALES DR	1,484
BRONSON AVE	1,484
MONTREAL RD	1,484
REGIONAL ROAD 174	1,484

Street Name. Color shows count of fact\_table. Size shows count of fact\_table. The marks are labeled by Street Name. The view is filtered on Street Name, which keeps 10 of 2,001 members.

Count of fact\_table

119 1,484

28







## HOW TO CONNECT PSQL TO PYTHON

- Create .ini file with all the details about server, database, user and password.
- Connecting to the database with the configuration file and psycopg2 package.
- We can now run SQL commands on python
- Full tutorial on:  
<https://www.postgresqltutorial.com/postgresql-python/connect/>

31

## USING SCIKIT LEARN

-  Scikit-learn is an awesome package for Machine learning.
-  Has a variety of models from decision trees to neural networks.
-  Each model needs to be fitted with training data, then tested with testing data.
-  More on that on next slides.

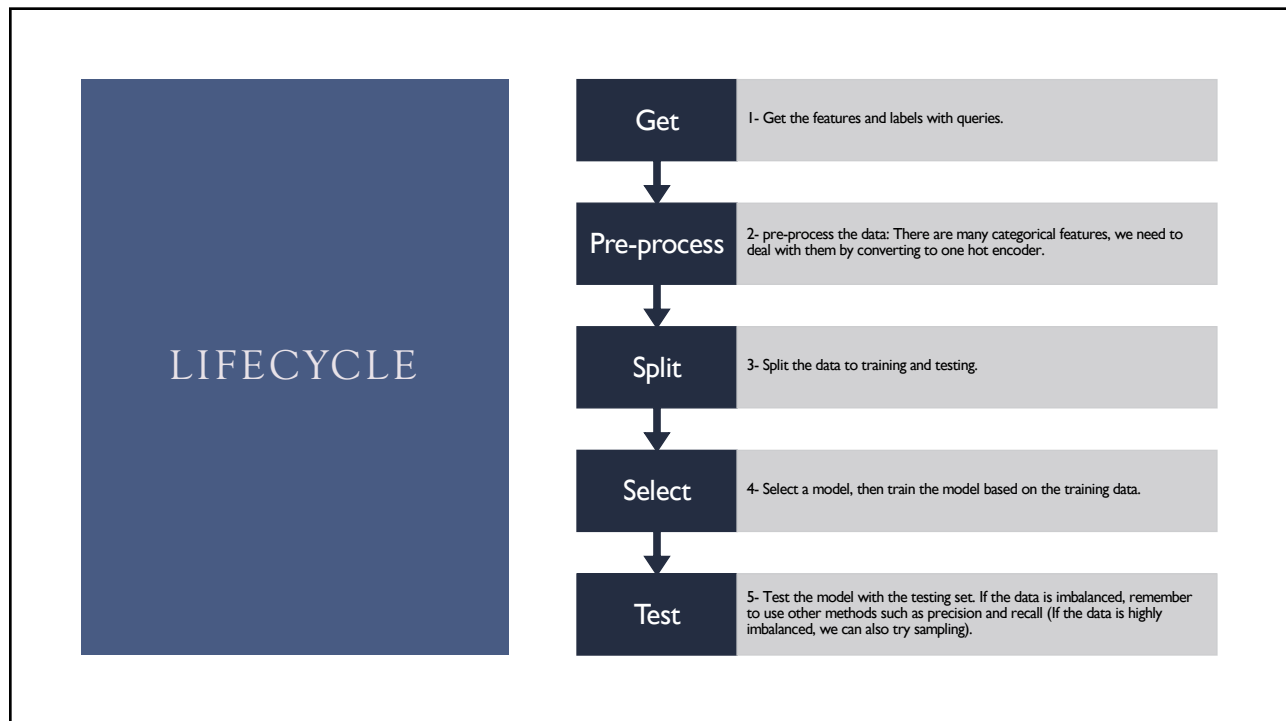
32



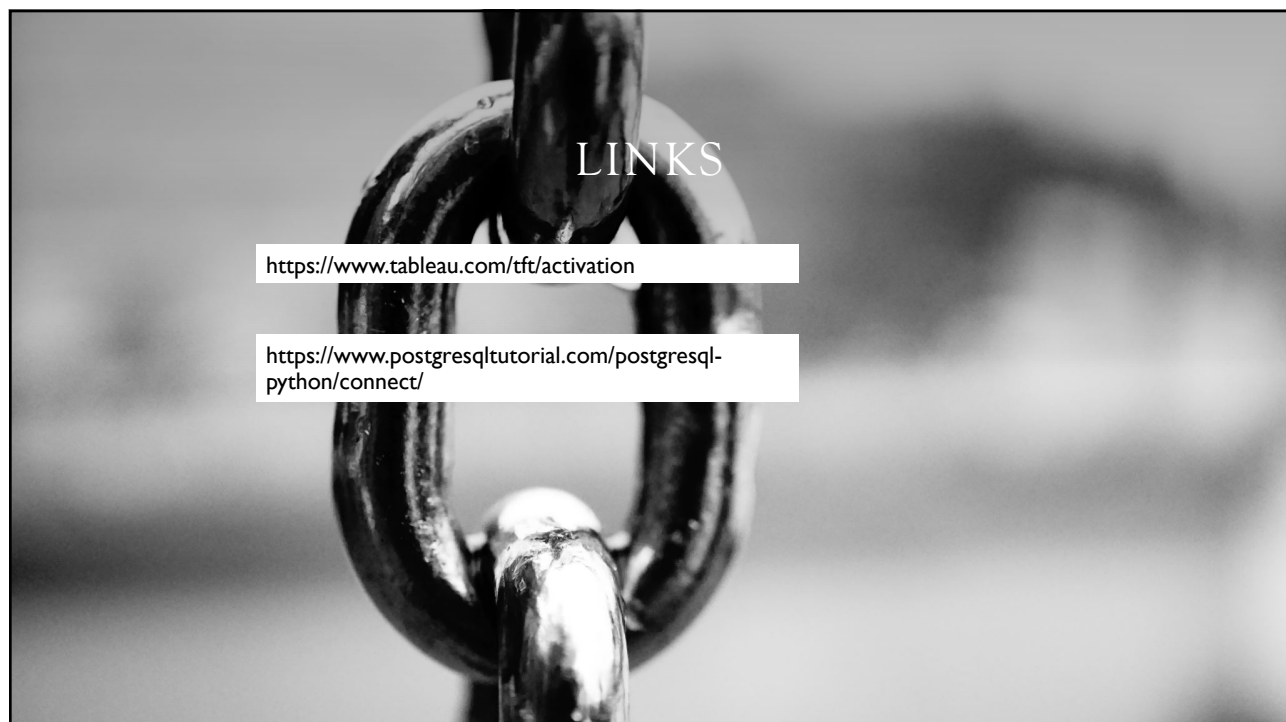
We need to see which features we want to use in our data that best describes the label.  
(the features)

34

weather	
weather_key	int
weather_ID	int
station_name	varchar
local_date	date
mean_temp	float
min_temp	float
max_temp	float
total_rain	float
total_snow	float
wind_speed	float



35



36