Fundamentals of Data Science

Project Phase 3: Data Analysis

By Johan Fernandes and Parsa Vafaie

jfern090@uottawa.ca and pvafa014@uottawa.ca

Overview of lecture

1. Download and Connect Tableau to PSQL
2. Revisiting the Accident data mart – 2019 project
3. Demo on OLAP queries:
    1. Drill Down
    2. Slice
    3. Dice
    4. Roll up
    5. Cube
    6. Window
    7. Iceberg

Download and Connect Tableau to PSQL instance

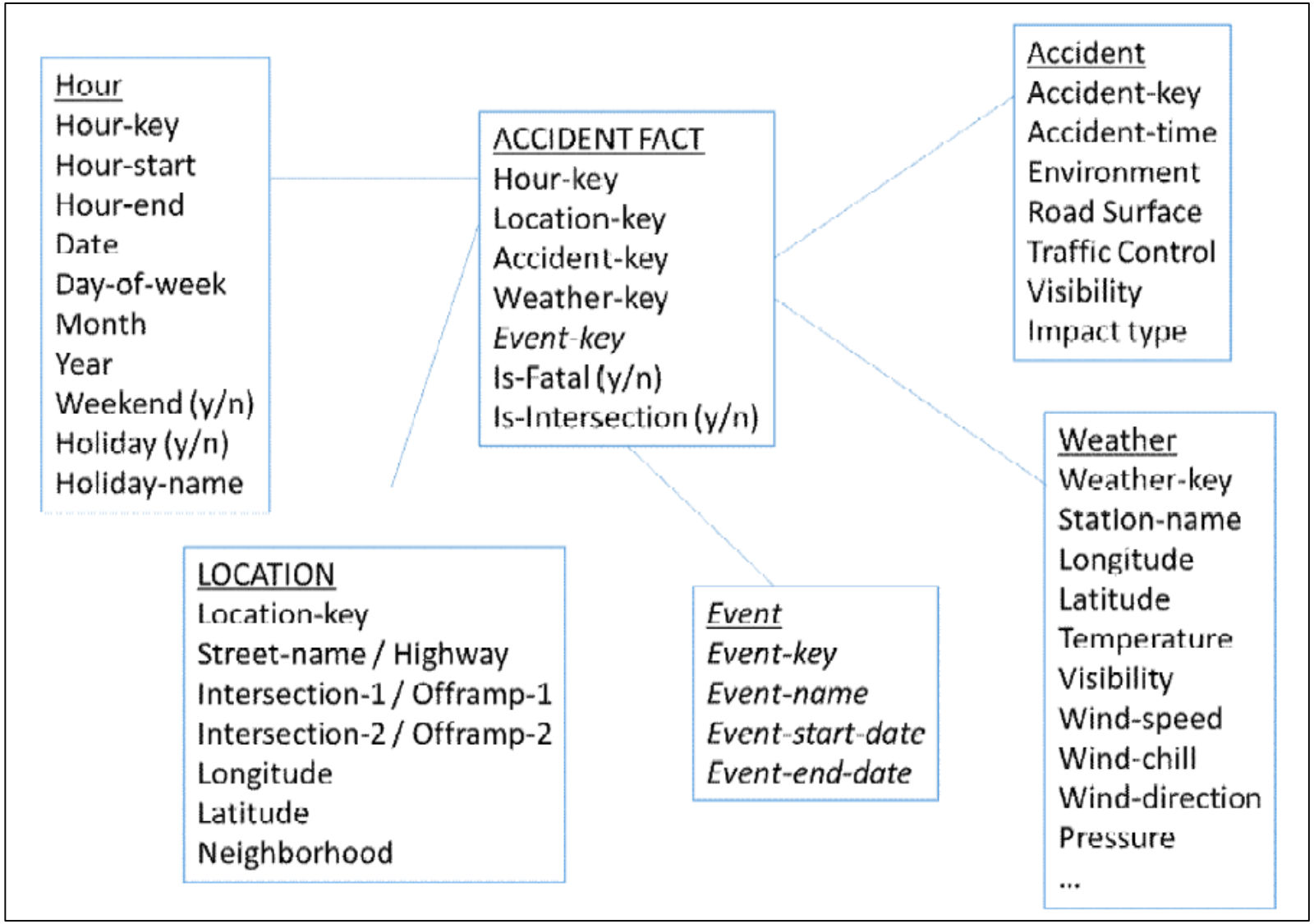
1. Download Tableau Desktop from this link: https://www.tableau.com/academic/students
2. Register yourself with your uOttawa id to ensure you get verified properly and avail the 1 year license.
3. Once Tableau Desktop is installed under *Connect* click on “To a Server” and select PostgreSQL
4. Enter the following details:
    1. Server: web0.eecs.uottawa.ca
    2. Port: 15432
    3. Database: group_<Your group number>
    4. Authentication: Username and Password. Note Username is the your uOttawa id i.e. the part before the @ symbol in your email address
    5. Select the Requires SSL checkbox.

    That’s it!

Revisiting Accident data mart

Source code available at:

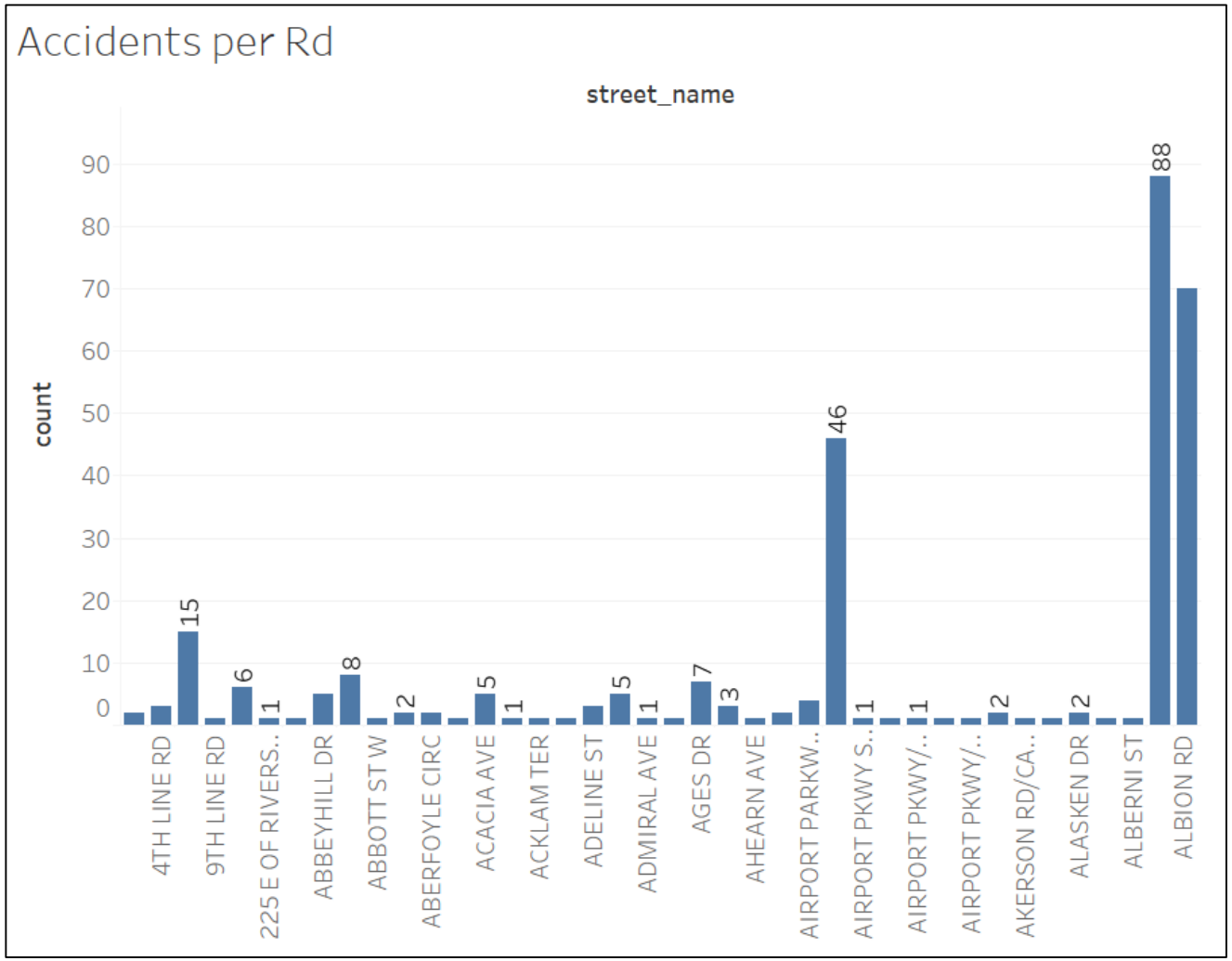
https://drive.google.com/drive/folders/1xuLCc-eD6ROqEbELWI3AevfuPW20XsYy?usp=sharing



Simple Query

Query Result: Number of accidents per road in 2017

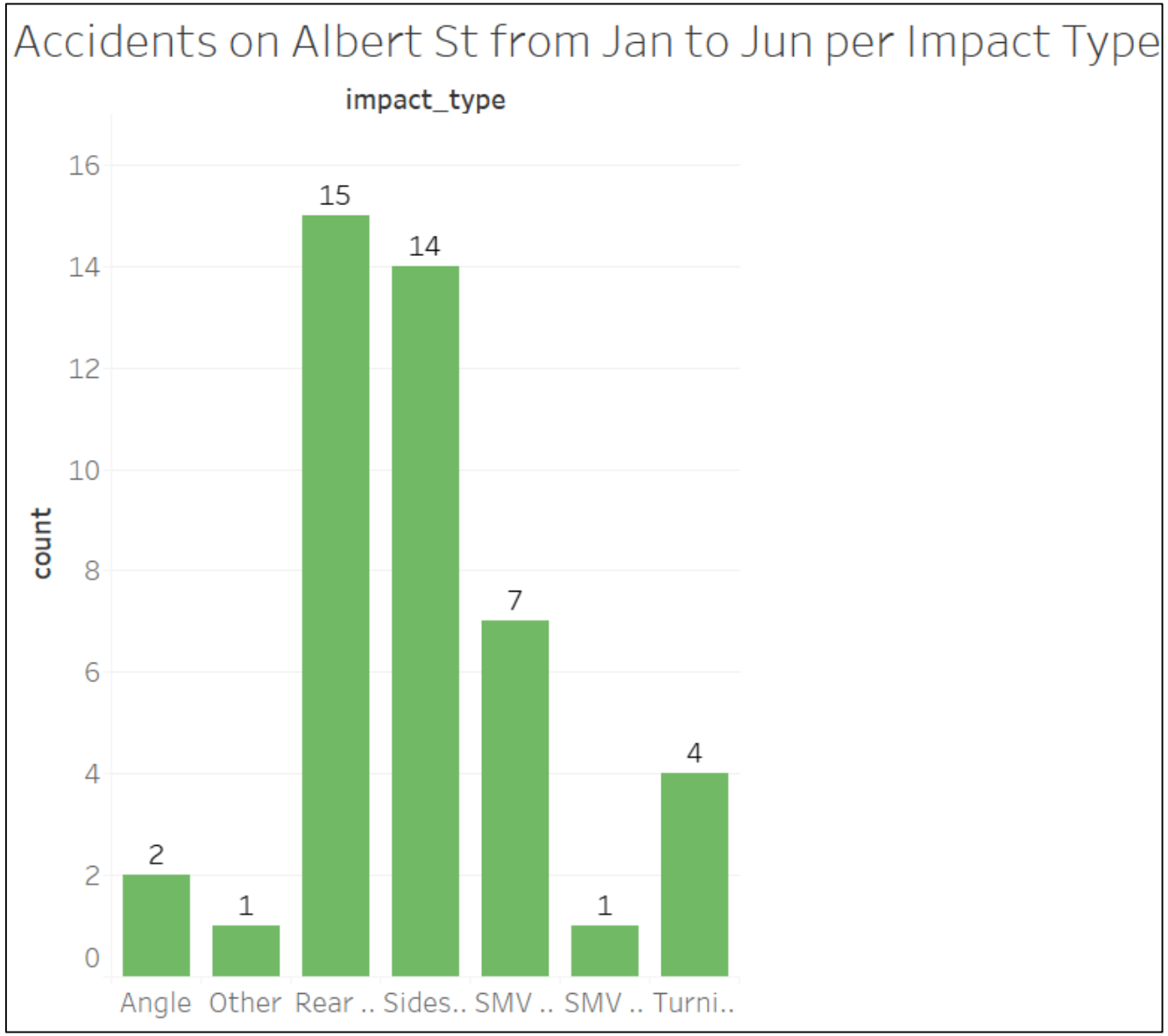
SELECT l.street_name, count(*) FROM location_dimension as l INNER JOIN accident_fact as a ON l.location_key = a.location_key GROUP BY street_name



Drill Down

Query Result: Number of accidents that occurred on Albert Street from January to June per type of impact

```
SELECT ad.impact_type, count(*) FROM
location_dimension as l INNER JOIN
accident_fact as a
ON l.location_key = a.location_key
INNER JOIN accident_dimension as ad
ON ad.accident_key = a.accident_key
INNER JOIN hour_dimension as h
ON h.hour_key = a.hour_key
WHERE street_name = 'ALBERT ST' AND
month BETWEEN 1 AND 6
GROUP BY impact_type
```



Slice

Query Result: Number of accidents that occurred based on type of impact, street name and month

SELECT ad.impact_type, l.street_name, h.month, count(*) FROM

location_dimension as l INNER JOIN

accident_fact as a

ON l.location_key = a.location_key

INNER JOIN accident_dimension as ad

ON ad.accident_key = a.accident_key

INNER JOIN hour_dimension as h

ON h.hour_key = a.hour_key

GROUP BY impact_type, street_name, month

Accidents per impact type per street for Feb 2017

		street_name									
month	impact_type	ABER..	ACACI..	ACKLA..	ADELI..	ADMI..	AIRPO..	AIRPO..	AIRPO..	ALAD..	ALBER..	ALBIO..
2	Angle				1					1		2
	Approaching		1					1				
	Other							1				
	Rear end							2	1		5	2
	Sideswipe										2	1
	SMV other										2	1
	SMV unattended..	1		1	2	1						
	Turning moveme..						1					1

Note: the Filter for month was added on Tableau rather than the SQL query

Dice

Query Result: Number of accidents that occurred based on type of impact, street name and month

SELECT ad.impact_type, l.street_name, h.month, count(*) FROM

location_dimension as l INNER JOIN

accident_fact as a

ON l.location_key = a.location_key

INNER JOIN accident_dimension as ad

ON ad.accident_key = a.accident_key

INNER JOIN hour_dimension as h

ON h.hour_key = a.hour_key

GROUP BY impact_type, street_name, month

Accidents per impact type for selected streets for Feb 2017

impact_type	month	street_name ADELINE ST	ALBERT ST	ALBION RD
Angle	2	1		2
Rear end	2		5	2
Sideswipe	2		2	1
SMV other	2		2	1
SMV unattended vehicle	2	2		
Turning movement	2			1

Note: the Filter for month and street names was added on Tableau rather than the SQL query

Roll Up

Query Result: Number of accidents grouped by impact type and street name as a group and individually

SELECT ad.impact_type, l.street_name,

GROUPING(ad.impact_type, l.street_name), SUM(a.is_intersection)

FROM accident_dimension as ad INNER JOIN

accident_fact as a

ON ad.accident_key = a.accident_key

INNER JOIN location_dimension as l

ON l.location_key = a.location_key

GROUP BY ROLLUP(impact_type, street_name)

ORDER BY impact_type, street_name

Acccidents per impact type and street

	impact_type								
street_name	Null	Angle	Appro..	Other	Rear e..	Sides..	SMV o..	SMV u..	Turnin..
210 W OF MERIV..					0				2
225 E OF RIVERS..					0				
2ND LINE RD							1		
4TH LINE RD							0		
8TH LINE RD		3		0	0		1		1
9TH LINE RD		1							
AARON AVE								0	
ABBEYHILL DR		1		0	0				
ABBOTT ST		2		0	0		1		
ABBOTT ST W					0				
ABERDEEN ST		1						0	
ABERFOYLE CIRC								0	
ABETTI RIDGE								0	
ACACIA AVE		1	0				0	0	1
ACCEPTANCE PL							0		
ACKLAM TER								0	
ADAMSON CRES								0	
ADELINE ST		1						1	
ADIRONDACK DR					1				
ADMIRAL AVE								0	
AERO DR						1			
AGES DR		0			1	0	0		0
AGINCOURT RD		0			0			0	
AHEARN AVE								0	

Cube

Query Result: Number of accidents grouped by impact type and street name as a group and individually

SELECT ad.impact_type, l.street_name,

GROUPING(ad.impact_type, l.street_name), SUM(a.is_intersection)

FROM accident_dimension as ad INNER JOIN

accident_fact as a

ON ad.accident_key = a.accident_key

INNER JOIN location_dimension as l

ON l.location_key = a.location_key

GROUP BY CUBE(impact_type, street_name)

ORDER BY impact_type, street_name

Acccidents per impact type and street

	impact_type								
street_name	Null	Angle	Approac..	Other	Rear end	Sideswipe	SMV other	SMV una..	Turning ..
210 W OF MERIV..	2				0				2
225 E OF RIVERS..	0				0				
2ND LINE RD	1						1		
4TH LINE RD	0						0		
8TH LINE RD	5	3		0	0		1		1
9TH LINE RD	1	1							
AARON AVE	0							0	
ABBEYHILL DR	1	1		0	0				
ABBOTT ST	3	2		0	0		1		
ABBOTT ST W	0				0				
ABERDEEN ST	1	1						0	
ABERFOYLE CIRC	0							0	
ABETTI RIDGE	0							0	
ACACIA AVE	2	1	0				0	0	1
ACCEPTANCE PL	0						0		
ACKLAM TER	0							0	
ADAMSON CRES	0							0	
ADELINE ST	2	1						1	
ADIRONDACK DR	1				1				
ADMIRAL AVE	0							0	
AERO DR	1					1			
AGES DR	1	0			1	0	0		0
AGINCOURT RD	0	0			0			0	
AHEARN AVE	0							0	

Window

Query Result: Number of accidents grouped by impact type and street name as a group and individually

SELECT h.month, w.temp, avg(w.temp) OVER W as avgtemp

FROM weather_dimension as w INNER JOIN

accident_fact as a

ON w.weather_key = a.weather_key

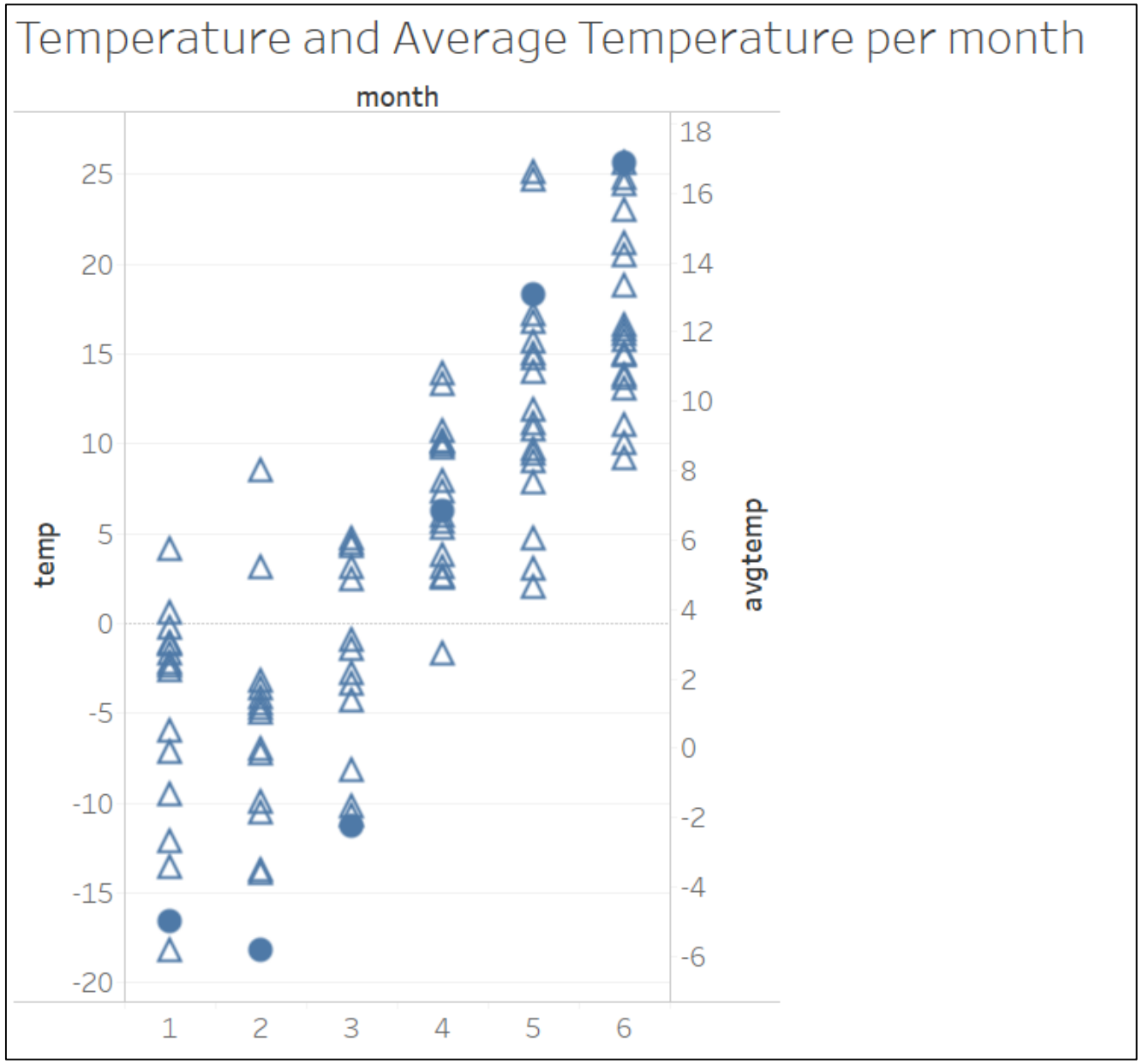
INNER JOIN hour_dimension as h

ON h.hour_key = a.hour_key

WHERE h.month BETWEEN 1 and 6

WINDOW W as (PARTITION by h.month

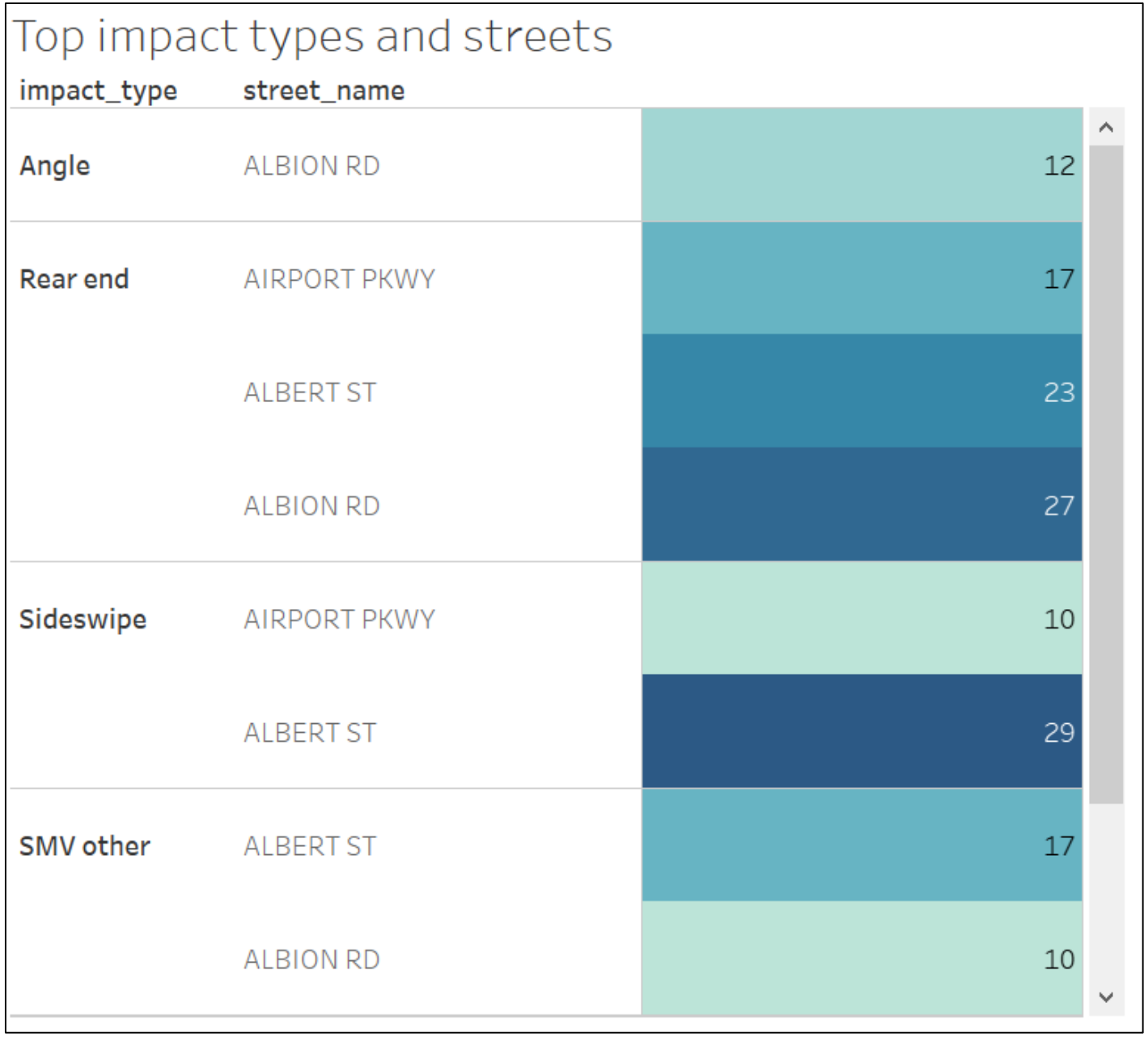
ORDER BY h.month)



Iceberg

Query Result: Top 10 accident counts grouped by impact type and street name.

```
SELECT ad.impact_type, l.street_name, count(*) as total FROM
location_dimension as l INNER JOIN
accident_fact as a
ON l.location_key = a.location_key
INNER JOIN accident_dimension as ad
ON ad.accident_key = a.accident_key
GROUP BY impact_type, street_name
ORDER BY total DESC
LIMIT 10
```



Resources:

For OLAP queries:

https://www.postgresqltutorial.com/postgresql-cube/

https://www.postgresqltutorial.com/postgresql-rollup/

https://www.postgresqltutorial.com/postgresql-window-function/

Refer to the official training videos on Tableau to learn how to use it efficiently for your project. You will have to create an account to watch the videos.

https://www.tableau.com/learn/training/20194

Questions?