

A Case Study

New York City Bike Sharing

Clemens S. Heithecker
and Thea A. Putnam



Photo by Anthony Fomin on Unsplash

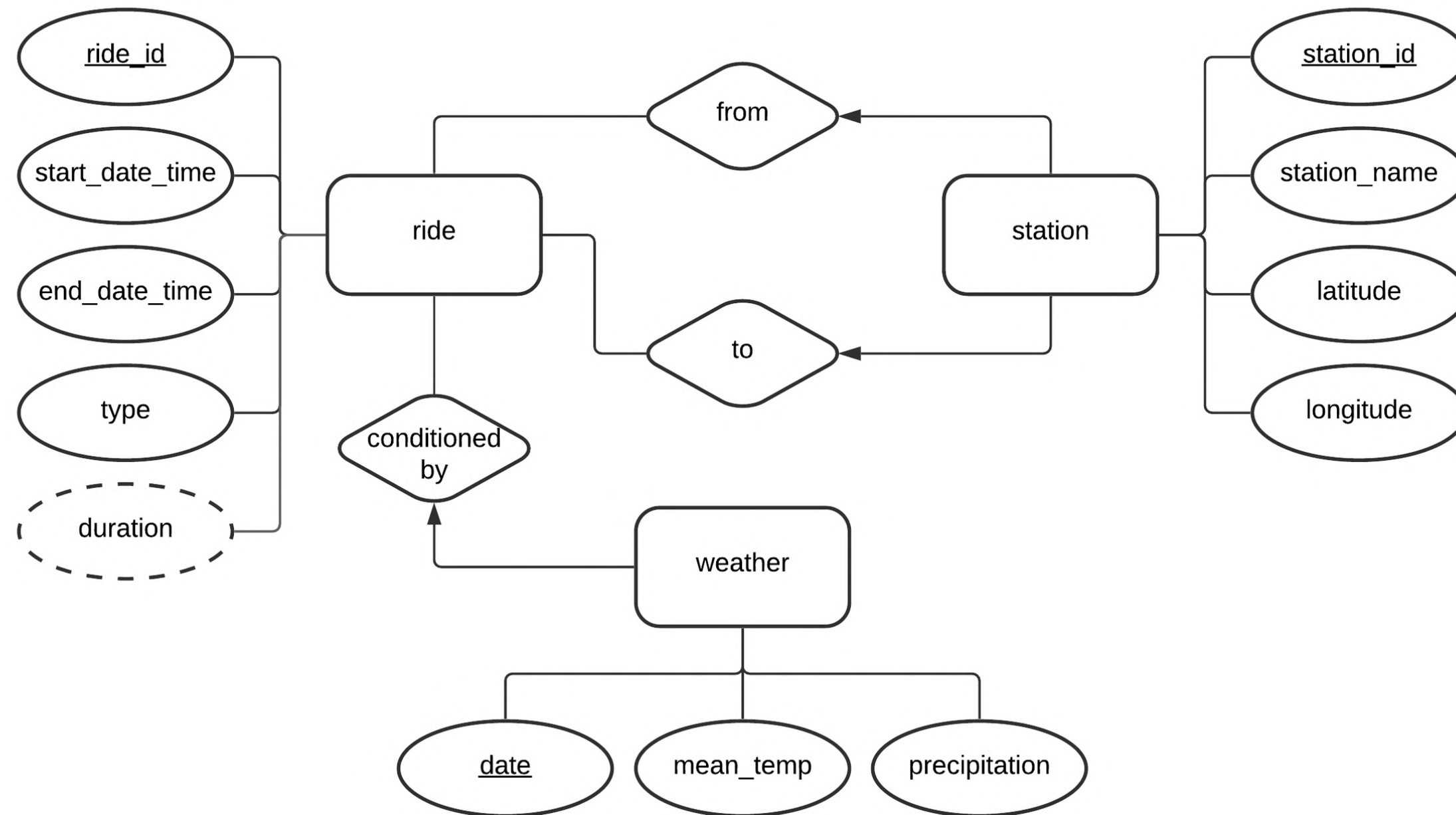
Datasets

JC-202110-citibike-tripdata.csv								NYC Bike Share, LLC & Jersey City Bike Share, LLC				
ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual
0FA0...	clasi...	2021-10-19 06:08:46	2021-10-19 06:11:26	Glenwo...	JC094	Sip Ave	JC056	40.7...	-74.1...	40.7...	-74.1...	member
0702...	clasi...	2021-10-...	2021-10-...	Glenwo...	JC094	Sip Ave	JC056	40.7...	-74.1...	40.7...	-74.1...	member
A881...	clasi...	2021-10-...	2021-10-...	Glenwo...	JC094	Sip Ave	JC056	40.7...	-74.1...	40.7...	-74.1...	member
C9A0...	clasi...	2021-10-...	2021-10-...	Glenwo...	JC094	Sip Ave	JC056	40.7...	-74.1...	40.7...	-74.1...	member
7DFD...	clasi...	2021-10-...	2021-10-...	Glenwo...	JC094	Sip Ave	JC056	40.7...	-74.1...	40.7...	-74.1...	member

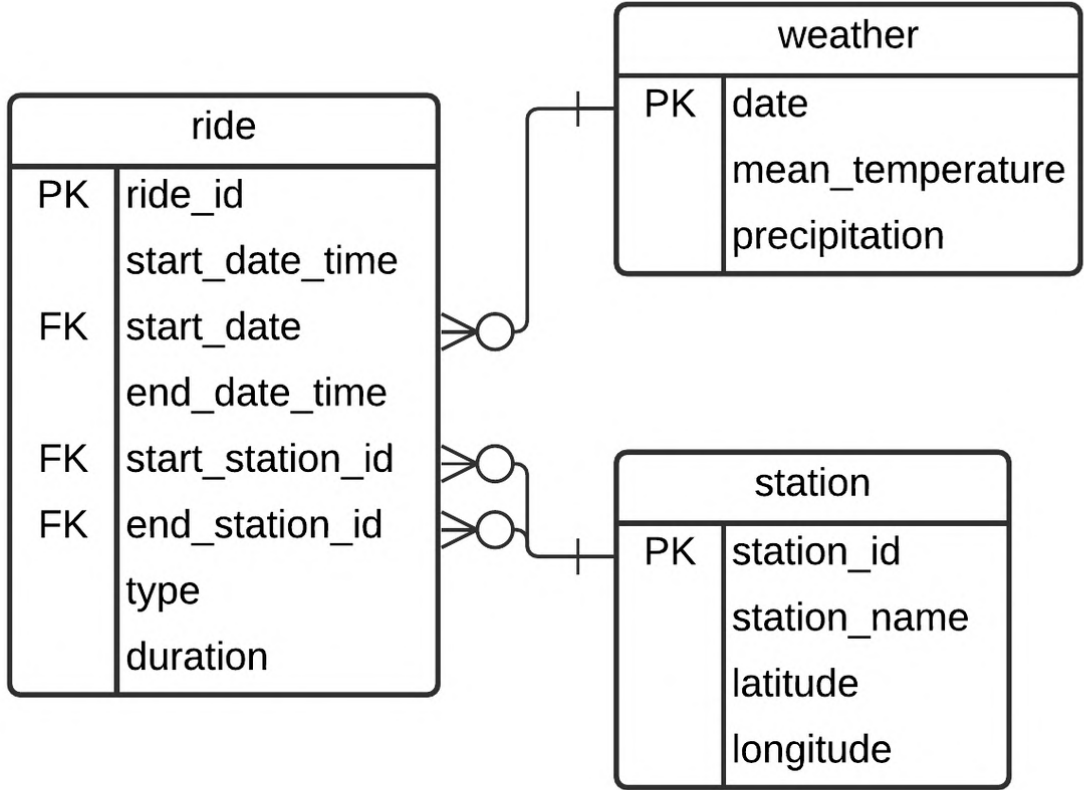
2796493.csv		NOAA National Climatic Data Center				
STATION	NAME	DATE	PRCP	TMAX	TMIN	
USW00094728	NY CITY CENTRAL PARK, NY US	2021-10-01	0	19.4	10.6	
USW00094728	NY CITY CENTRAL PARK, NY US	2021-10-02	0	23.3	13.9	
USW00094728	NY CITY CENTRAL PARK, NY US	2021-10-03	0	25	16.1	
USW00094728	NY CITY CENTRAL PARK, NY US	2021-10-04	3.8	22.8	16.7	
USW00094728	NY CITY CENTRAL PARK, NY US	2021-10-05	0	18.8	15.6	

New York City Bike Sharing Database

Entity Relationships (1)



New York City Bike Sharing Database
Entity Relationships (2)



Building and using the database

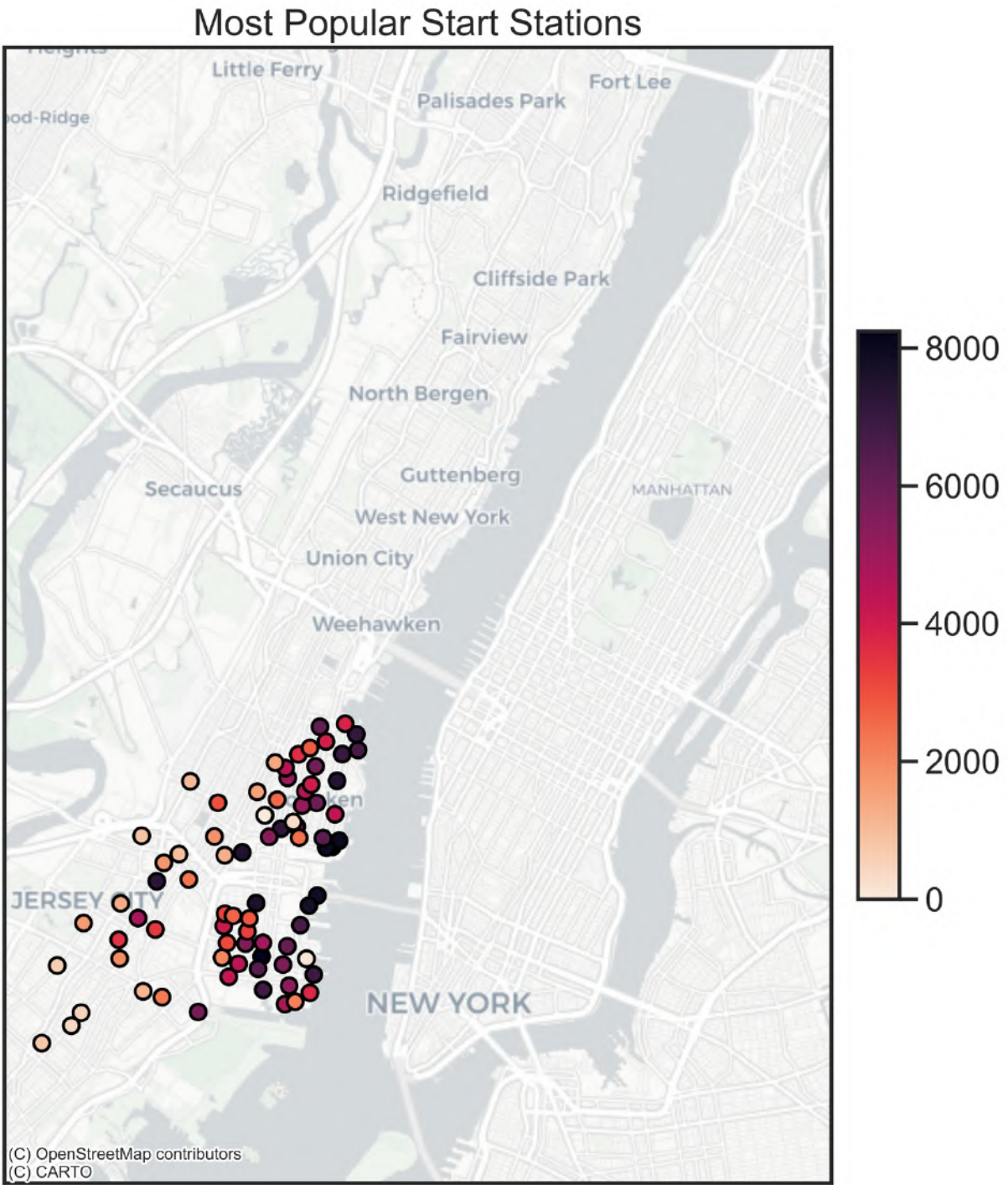
- 01 **Clean data**
- 02 **Create database schema**
- 03 **Import data**
- 04 **Query database**

NYC Bike Sharing

Station Analysis

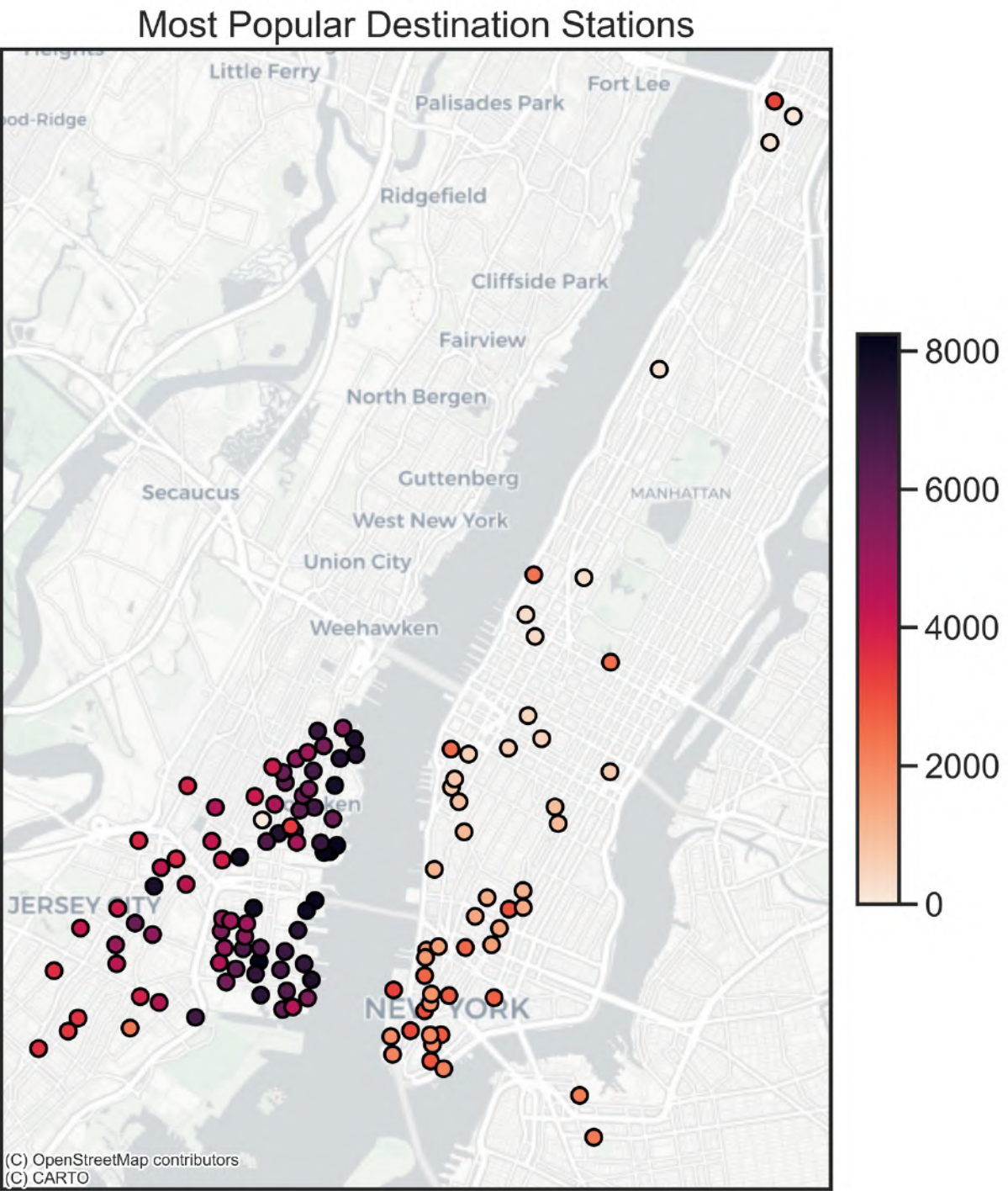
What are the *most popular* start stations?

station_id	station_name	latitude	longitude	ride_count
JC005	Grove St ...	40.7195861	-74.043117	4084
HB102	Hoboken T...	40.7360677	-74.029127	3687
HB103	South Wat...	40.7369822	-74.027781	3411
HB101	Hoboken T...	40.7359376	-74.030305	3203
JC008	Newport Pkwy	40.7287448	-74.032108	2445
JC066	Newport PATH	40.7272235	-74.033759	2318
JC009	Hamilton ...	40.7275960	-74.044247	2255



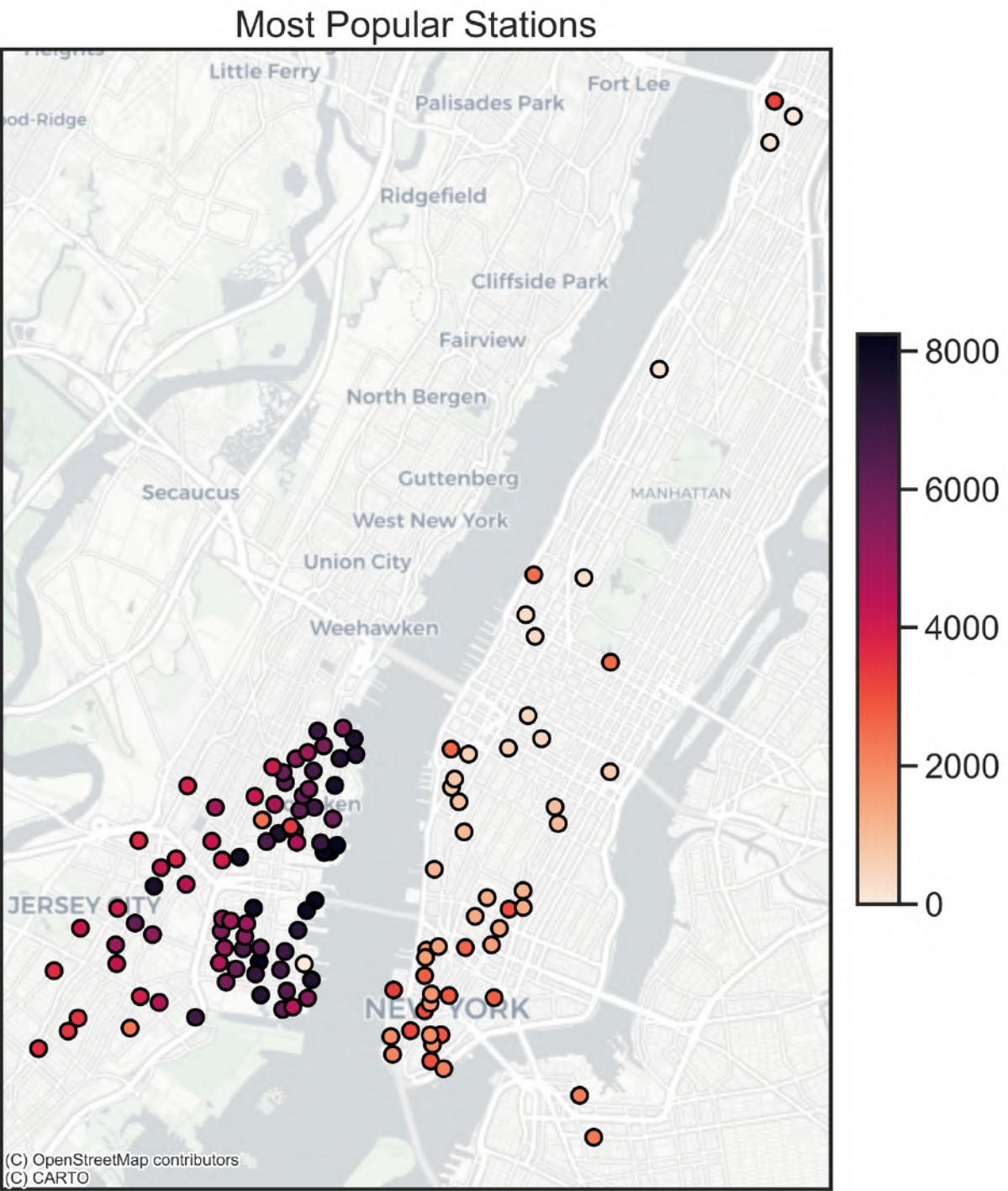
What are the *most popular* destination stations?

station_id	station_name	latitude	longitude	ride_count
JC005	Grove St ...	40.7195861	-74.043117	4156
HB102	Hoboken T...	40.7360677	-74.029127	3679
HB103	South Wat...	40.7369822	-74.027781	3447
HB101	Hoboken T...	40.7359376	-74.030305	3242
JC008	Newport Pkwy	40.7287448	-74.032108	2446
JC066	Newport PATH	40.7272235	-74.033759	2384
JC009	Hamilton ...	40.7275960	-74.044247	2276



What are the *most popular* stations overall?

station_id	station_name	latitude	longitude	ride_count
JC005	Grove St ...	40.7195861	-74.043117	8240
HB102	Hoboken T...	40.7360677	-74.029127	7366
HB103	South Wat...	40.7369822	-74.027781	6858
HB101	Hoboken T...	40.7359376	-74.030305	6445
JC008	Newport Pkwy	40.7287448	-74.032108	4891
JC066	Newport PATH	40.7272235	-74.033759	4702
JC009	Hamilton ...	40.7275960	-74.044247	4531

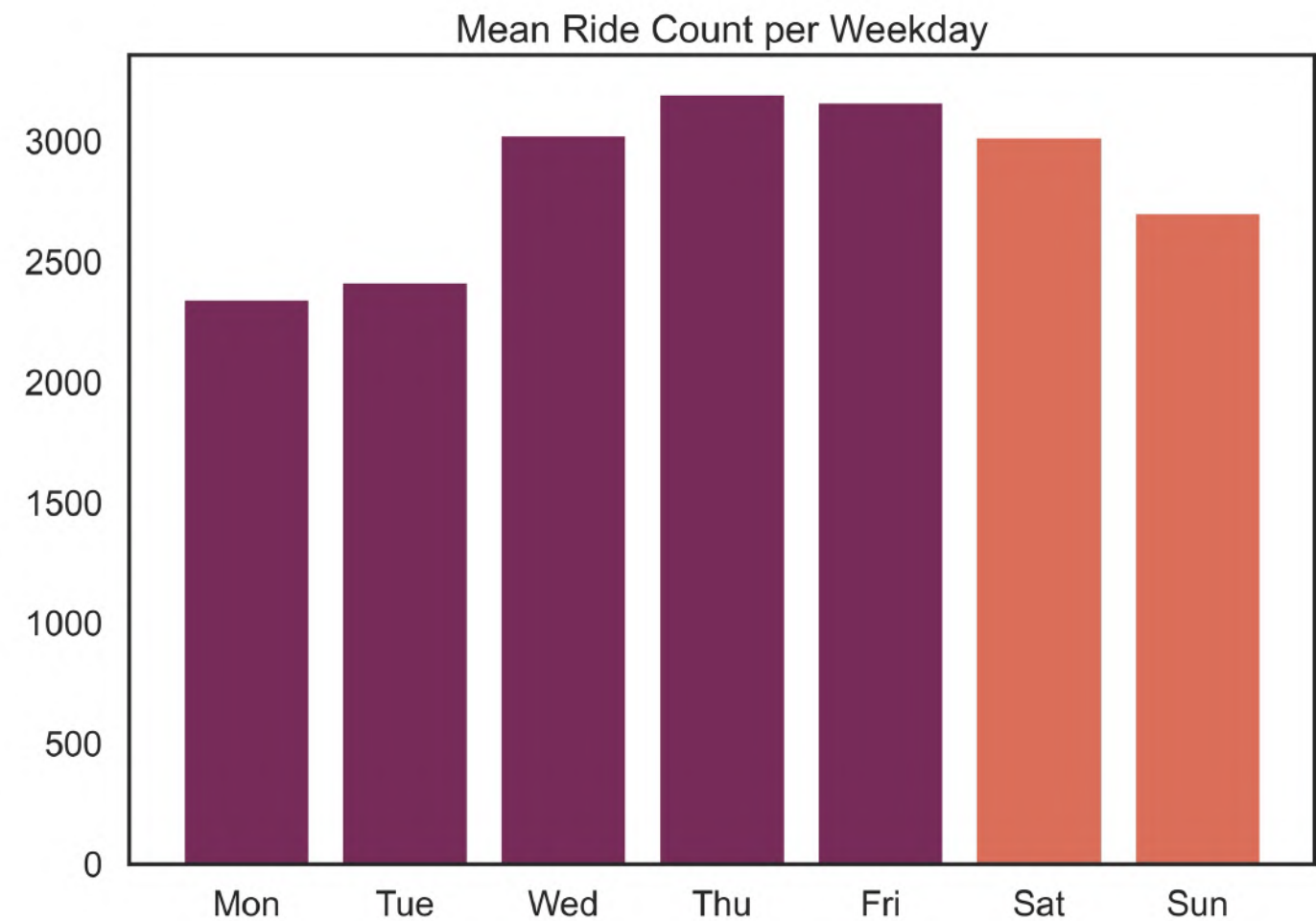


NYC Bike Sharing

Ride Analysis

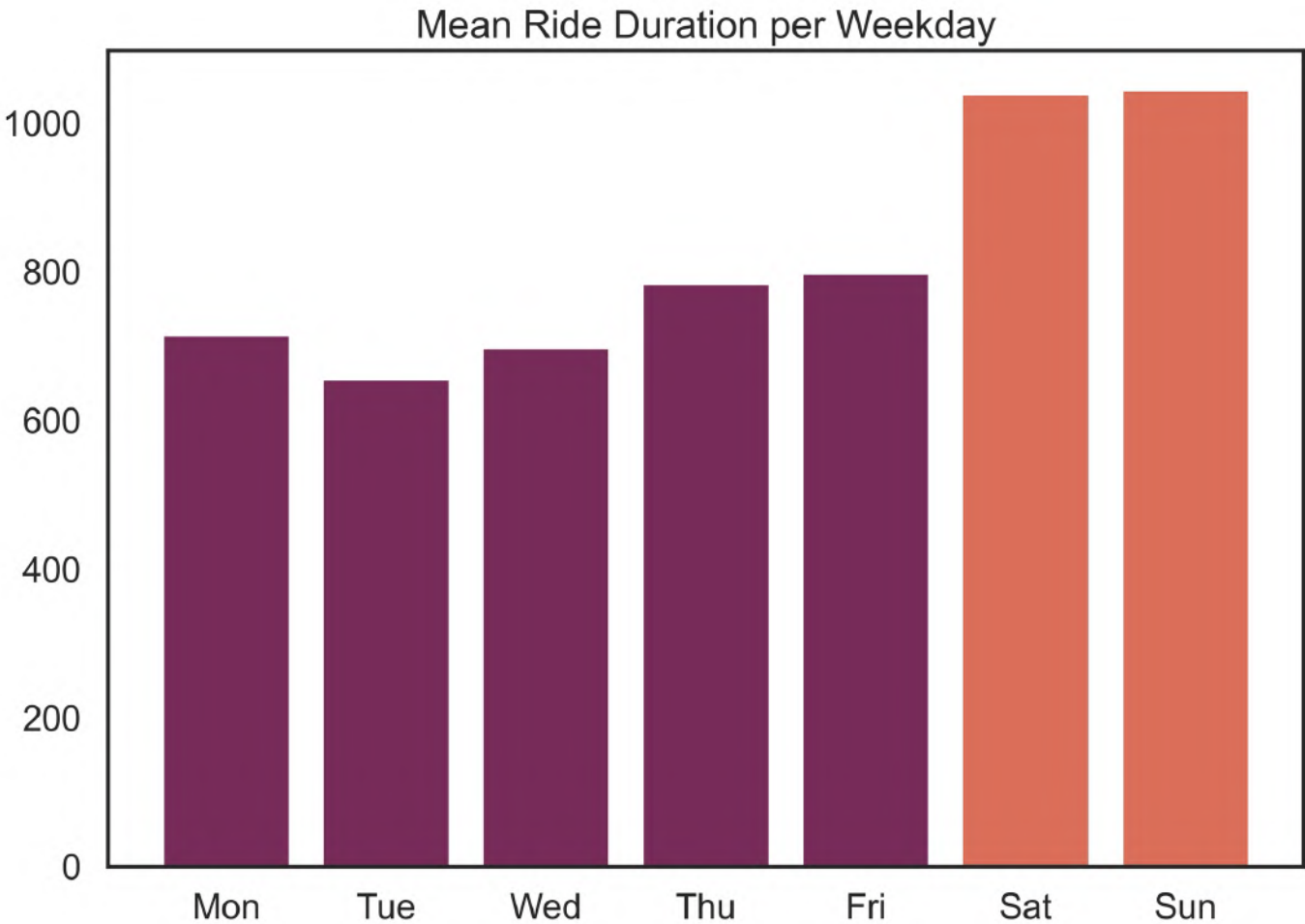
Do people rent bikes *more often* on weekends?

weekday	mean_ride_count
3	3198.00
4	3162.60
2	3026.25
5	3018.80
6	2705.60
1	2416.75
0	2344.25



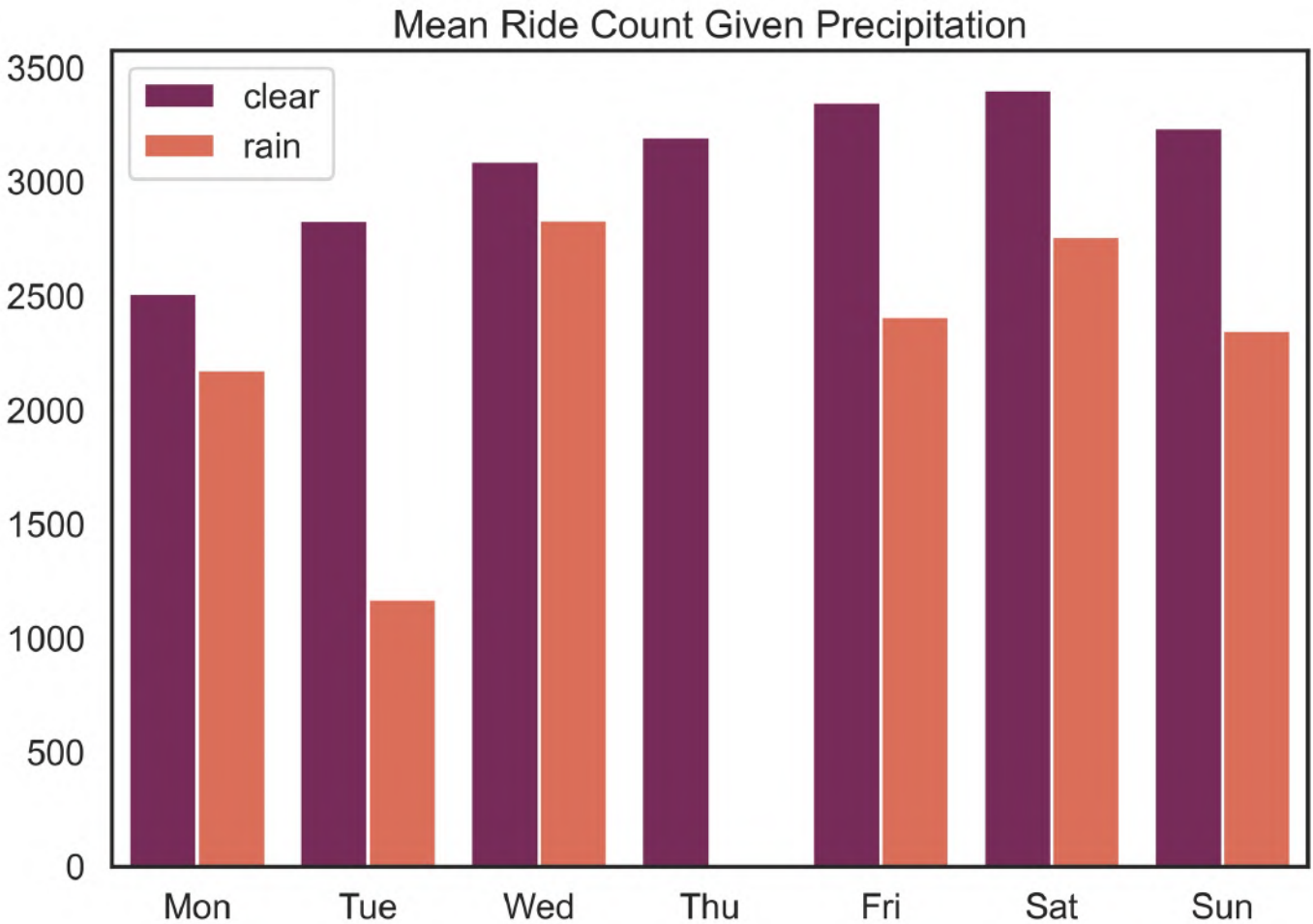
Do people rent bikes *for longer* on weekends?

weekday	mean_duration
6	1045.40228
5	1039.62846
4	798.78530
3	784.24461
0	715.66428
2	698.48600
1	656.49385



Do people rent bikes *less often* on rainy days?

weekday	mean Ride count _rain	mean Ride count _clear	mean Ride count _difference
0	2177.00	2511.50	-334.50
1	1173.00	2831.33	-1658.33
2	2834.00	3090.33	-256.33
3	NULL	3196.00	NULL
4	2412.00	3350.25	-938.25
5	2761.00	3405.50	-644.50
6	2351.00	3237.50	-886.50





Check out our project on GitHub

Clemens S. Heithecker

github.com/clemensheithecker/nyc-bike-sharing

Thea A. Putnam

github.com/theaputnam/nyc-bike-sharing

Thank you!

NYC Bike Sharing

SQL Queries

What are the *most popular* start stations?

```
CREATE VIEW TopStartStations AS
```

```
WITH cte_start_station_ride_count AS (  
    SELECT start_station_id, COUNT(ride_id) AS  
ride_count  
    FROM ride  
    GROUP BY start_station_id  
)
```

```
SELECT station_id, station_name, latitude,  
longitude, ride_count  
FROM (  
    station AS s  
    JOIN cte_start_station_ride_count AS c  
    ON s.station_id = c.start_station_id  
)  
ORDER BY ride_count DESC;
```

```
SELECT *  
FROM TopStartStations;
```

What are the *most popular* stations overall?

```
CREATE VIEW TopStations AS
WITH cte_start_station_ride_count AS (
    SELECT start_station_id, COUNT(ride_id) AS
ride_count
    FROM ride
    GROUP BY start_station_id
), cte_end_station_ride_count AS (
    SELECT end_station_id, COUNT(ride_id) AS
ride_count
    FROM ride
    GROUP BY end_station_id
), cte_total_station_ride_count AS (
    SELECT s.start_station_id AS station_id,
s.ride_count+e.ride_count AS ride_count
    FROM (
        SELECT *
        FROM cte_start_station_ride_count) s
    LEFT JOIN (
        SELECT *
        FROM cte_end_station_ride_count) e
    ON s.start_station_id = e.end_station_id
    UNION ALL
```

```
    SELECT e.end_station_id AS station_id,
e.ride_count
    FROM (
        SELECT *
        FROM cte_start_station_ride_count) s
    RIGHT JOIN (
        SELECT *
        FROM cte_end_station_ride_count) e
    ON s.start_station_id = e.end_station_id
    WHERE s.start_station_id IS NULL
)
SELECT s.station_id, s.station_name, s.latitude,
s.longitude, c.ride_count
FROM (
    station AS s
    JOIN cte_total_station_ride_count AS c
    ON s.station_id = c.station_id
)
ORDER BY c.ride_count DESC;

FROM TopStations;
SELECT *
```

Do people rent bikes *more often* on weekends?

```
CREATE VIEW RideCountPerWeekday AS
```

```
SELECT days.weekday, AVG(ride_count) AS  
mean_ride_count  
FROM (  
    SELECT start_date, WEEKDAY(start_date) AS  
weekday, COUNT(ride_id) AS ride_count  
    FROM ride  
    GROUP BY start_date  
) AS days  
GROUP BY days.weekday  
ORDER BY mean_ride_count DESC;
```

```
SELECT *  
FROM RideCountPerWeekday;
```


Do people rent bikes *for longer* on weekends?

```
CREATE VIEW RideDurationPerWeekday AS
```

```
SELECT WEEKDAY(start_date) AS weekday,  
AVG(duration) AS mean_duration  
FROM ride  
GROUP BY weekday  
ORDER BY mean_duration DESC;
```

```
SELECT *  
FROM RideDurationPerWeekday;
```

Do people rent bikes *less often* on rainy days?

```
CREATE VIEW PrecipitationEffect AS
WITH cte_ride_count_rain AS (
    SELECT days.weekday, AVG(days.ride_count) AS
mean_ride_count, 1 AS precipitation
    FROM (
        SELECT ride.start_date,
WEEKDAY(ride.start_date) AS weekday,
weather.precipitation, COUNT(ride.ride_id) AS
ride_count
        FROM ride LEFT JOIN weather ON
ride.start_date = weather.date
        GROUP BY ride.start_date
        HAVING weather.precipitation > 0
    ) AS days
    GROUP BY days.weekday
), cte_ride_count_clear AS (
    SELECT days.weekday, AVG(days.ride_count) AS
mean_ride_count, 0 AS precipitation
    FROM (
        SELECT WEEKDAY(ride.start_date) AS
weekday, weather.precipitation,
COUNT(ride.ride_id) AS ride_count
```

```
    FROM ride LEFT JOIN weather ON
ride.start_date = weather.date
    GROUP BY ride.start_date
    HAVING weather.precipitation = 0
    ) AS days
    GROUP BY days.weekday
)
```

```
SELECT rain.weekday, rain.mean_ride_count AS
mean_ride_count_rain, clear.mean_ride_count AS
mean_ride_count_clear, rain.mean_ride_count-
clear.mean_ride_count AS
mean_ride_count_difference
FROM (
    SELECT *
    FROM cte_ride_count_rain) rain
LEFT JOIN (
    SELECT *
    FROM cte_ride_count_clear) clear
ON rain.weekday = clear.weekday

UNION ALL
```

Do people rent bikes *less often* on rainy days?

-- continued...

```
SELECT clear.weekday, rain.mean_ride_count AS
mean_ride_count_rain, clear.mean_ride_count AS
mean_ride_count_clear, rain.mean_ride_count-
clear.mean_ride_count AS
mean_ride_count_difference
FROM (
    SELECT *
    FROM cte_ride_count_rain) rain
RIGHT JOIN (
    SELECT *
    FROM cte_ride_count_clear) clear
ON rain.weekday = clear.weekday
WHERE rain.weekday IS NULL
ORDER BY weekday;

SELECT *
FROM PrecipitationEffect;
```