

Weather's Effect on Traffic Incidents

In Portland and Seattle

Avery Pike & Tyler Gomez Riddick

A dark blue diagonal gradient bar that starts from the bottom left and extends towards the top right, covering the lower half of the slide.

The Data

- 2.5 Openweather API
 - Current Weather

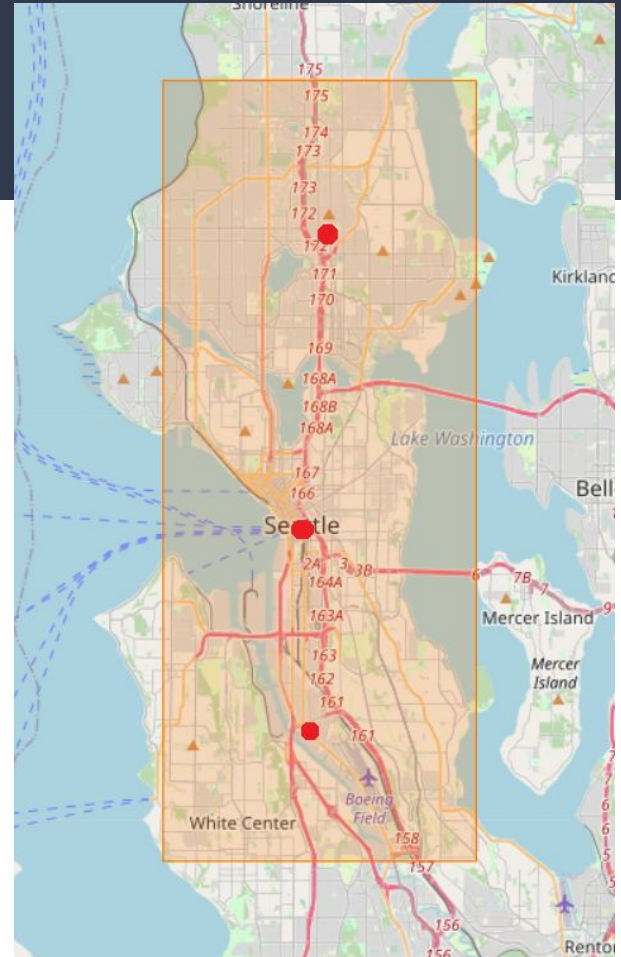
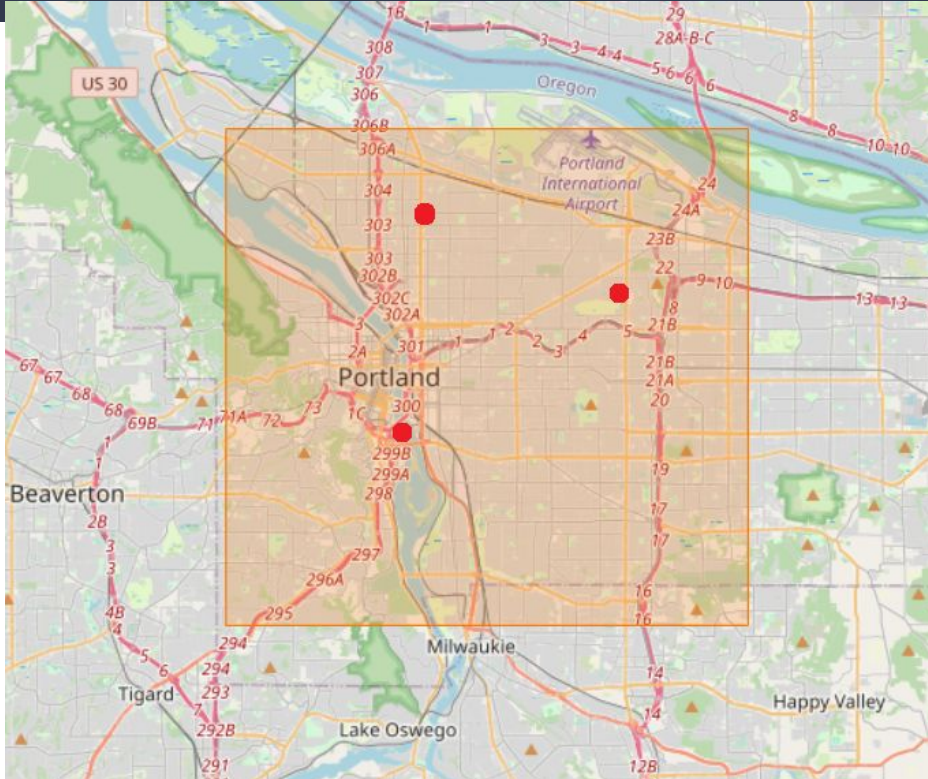


- Mapquest Traffic Incidents API



- Chose to look at Seattle and Portland
 - Large Cities
 - Heavy Precipitation
 - Variable Weather in the Spring

Maps



```
{
  "incidents": [
    {
      "id": "2273855287708804011",
      "type": 4,
      "severity": 3,
      "eventCode": 401,
      "lat": 47.51141,
      "lng": -122.29585,
      "startTime": "2024-04-24T13:19:34",
      "endTime": "2024-04-26T19:19:34",
      "impacting": true,
      "shortDesc": "Closed",
      "fullDesc": "Closed",
      "delayFromFreeFlow": 0,
      "delayFromTypical": 0,
      "distance": 0.12181479179614667,
      "iconURL": "http://content.mqcdn.com/mqtraffic/incid_sev.png",
      "parameterizedDescription": {}
    },
    {
      "id": "1064660022454052013",
      "type": 1,
      "severity": 1,
      "eventCode": 703,
      "lat": 47.50632,
      "lng": -122.25449,
      "startTime": "2022-11-28T08:00:00",
      "endTime": "2025-01-02T07:59:00",
      "impacting": true,
      "shortDesc": "Road maintenance operations",
      "fullDesc": "Between 74th Ave S and Martin Luther King Jr Way S - Road maintenance operations",
      "delayFromFreeFlow": 0,
      "delayFromTypical": 0,
      "distance": 2.4244872591671847,
      "iconURL": "http://content.mqcdn.com/mqtraffic/const_min.png",
      "parameterizedDescription": {}
    }
  ]
}
```

```
{
  "coord": {
    "lon": -122.3243,
    "lat": 47.6901
  },
  "weather": [
    {
      "id": 501,
      "main": "Rain",
      "description": "moderate rain",
      "icon": "10d"
    }
  ],
  "base": "stations",
  "main": {
    "temp": 282.43,
    "feels_like": 280.83,
    "temp_min": 282.43,
    "temp_max": 282.43,
    "pressure": 1007,
    "humidity": 94,
    "sea_level": 1007,
    "grnd_level": 998
  },
  "visibility": 10000,
  "wind": {
    "speed": 2.94,
    "deg": 128,
    "gust": 9.23
  },
  "rain": {
    "1h": 1.19
  },
  "clouds": {
    "all": 100
  },
  "dt": 1714091432,
  "sys": {
    "country": "US",
    "sunrise": 1714050049,
    "sunset": 1714101190
  },
  "timezone": -25200,
  "id": 5810301,
  "name": "Shoreline",
  "cod": 200
}
```

Raw json Data Formats

Data Processing and Obstacles

- Data in Weather API was nested in a way that made populating tables in 3rd normal form difficult
 - Used a cross join lateral statement to populate the station table
- Joined on pulled at and Station ID
 - Incidents had their own coordinates
 - Calculated distance to nearest station
 - Iterated through stations and created a new column in incidents for nearest station

seattle_weather_records

🔑 dt	timestamp
🔑 station_id	int
temp	float
feels_like	float
temp_min	float
temp_max	float
pressure	float
humidity	float
visibility	float
wind_speed	float
rain_hour	float
clouds	float
weather	text
description	text

station

🔑 id	bigint
lon	float
lat	float

portland_weather_records

🔑 dt	timestamp
🔑 station_id	int
temp	float
feels_like	float
temp_min	float
temp_max	float
pressure	float
humidity	float
visibility	float
wind_speed	float
rain_hour	float
clouds	float
weather	text
description	text

```

INSERT INTO weather_records
(dt, station_id, temp, feels_like, temp_min, temp_max, pressure,
 humidity, visibility, wind_speed, rain_hour, clouds, weather, description)
SELECT
to_timestamp((sr.raw_json->>'dt')::BIGINT) AS dt,
s.id AS station_id,
(sr.raw_json->'main'->>'temp')::NUMERIC AS temp,
(sr.raw_json->'main'->>'feels_like')::NUMERIC AS feels_like,
(sr.raw_json->'main'->>'temp_min')::NUMERIC AS temp_min,
(sr.raw_json->'main'->>'temp_max')::NUMERIC AS temp_max,
(sr.raw_json->'main'->>'pressure')::NUMERIC AS pressure,
(sr.raw_json->'main'->>'humidity')::NUMERIC AS humidity,
(sr.raw_json->>'visibility')::NUMERIC AS visibility,
(sr.raw_json->'wind'->>'speed')::NUMERIC AS wind_speed,
(sr.raw_json->'rain'->>'1h')::NUMERIC AS rain_hour,
(sr.raw_json->'clouds'->>'all')::NUMERIC AS clouds,
(w->>'main') AS weather,
(w->>'description') AS description
FROM
  seattle sr
JOIN
  station s ON (sr.raw_json->'coord'->>'lon')::NUMERIC = s.lon AND
  (sr.raw_json->'coord'->>'lat')::NUMERIC = s.lat
CROSS JOIN LATERAL
  jsonb_array_elements(sr.raw_json->'weather') AS w;

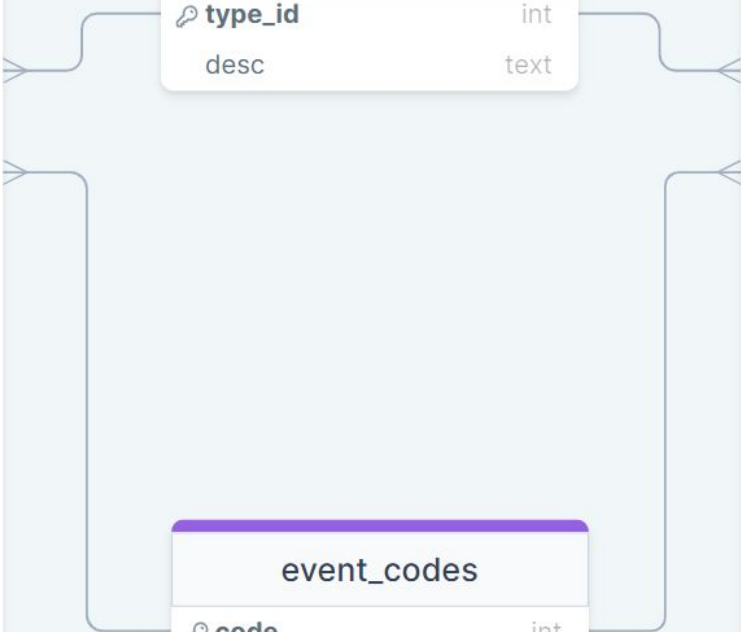
```

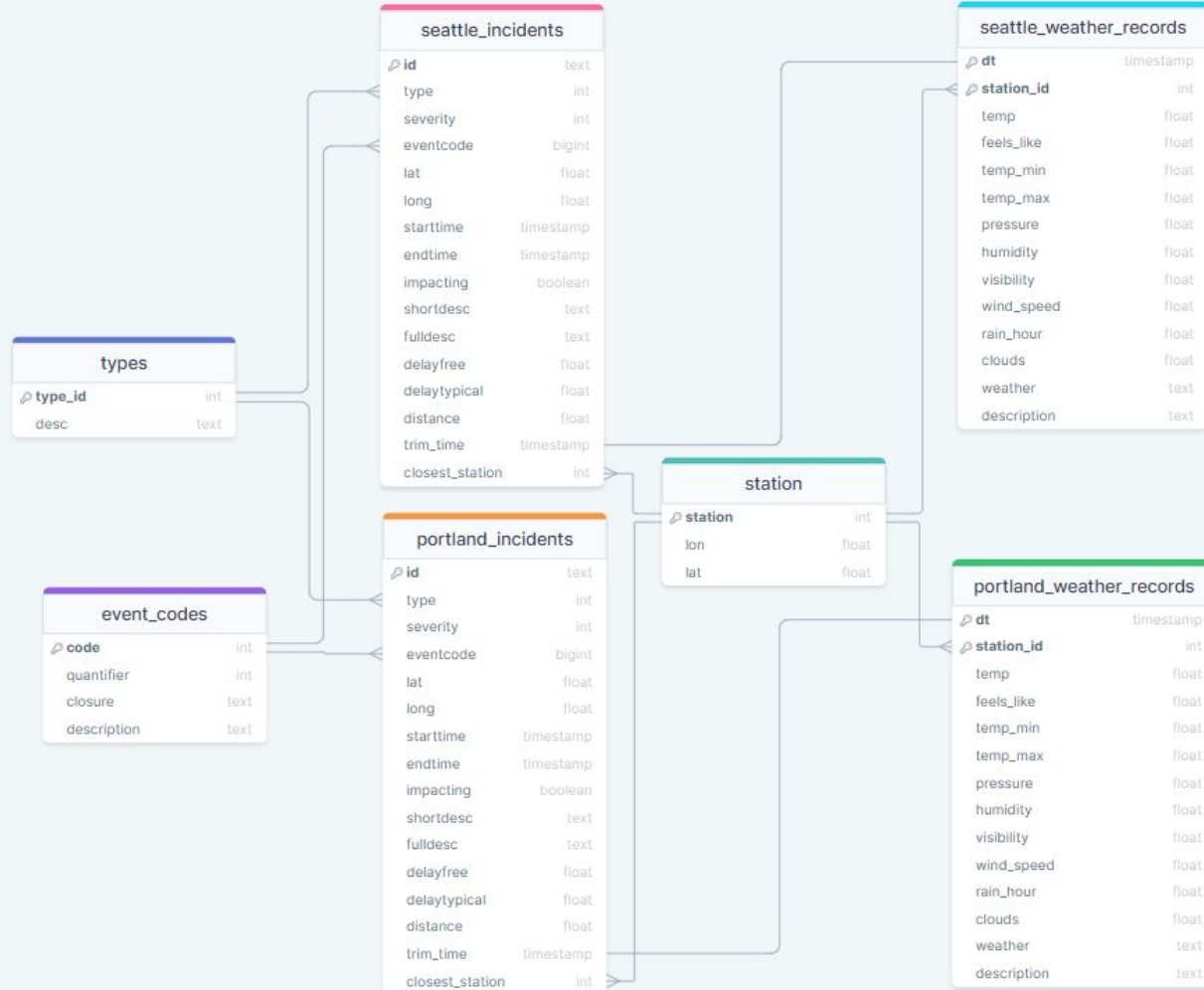

seattle_incidents	
🔑 id	text
type	int
severity	int
eventcode	bigint
lat	float
long	float
starttime	timestamp
endtime	timestamp
impacting	boolean
shortdesc	text
fulldesc	text
delayfree	float
delaytypical	float
distance	float
trim_time	timestamp
closest_station	int

types	
🔑 type_id	int
desc	text

portland_incidents	
🔑 id	text
type	int
severity	int
eventcode	bigint
lat	float
long	float
starttime	timestamp
endtime	timestamp
impacting	boolean
shortdesc	text
fulldesc	text
delayfree	float
delaytypical	float
distance	float
trim_time	timestamp
closest_station	int

event_codes	
🔑 code	int
quantifier	int
closure	text
description	text





Our API

- Endpoints:
 - <https://pikegomezapi-production.up.railway.app/cityweather>
 - <https://pikegomezapi-production.up.railway.app/portland>
 - <https://pikegomezapi-production.up.railway.app/seattle>
- Outputs:
 - Weather Report at the stations within the city:
 - Temperature
 - Weather at the time
 - Count of Accidents
 - Count of Congestion Incidents

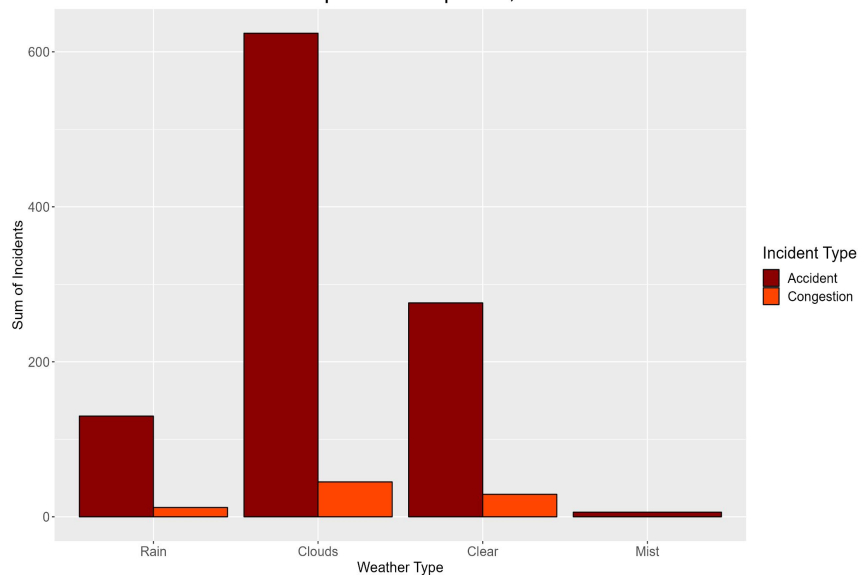
Does rain cause more accidents?

- Hypothesis: Yes

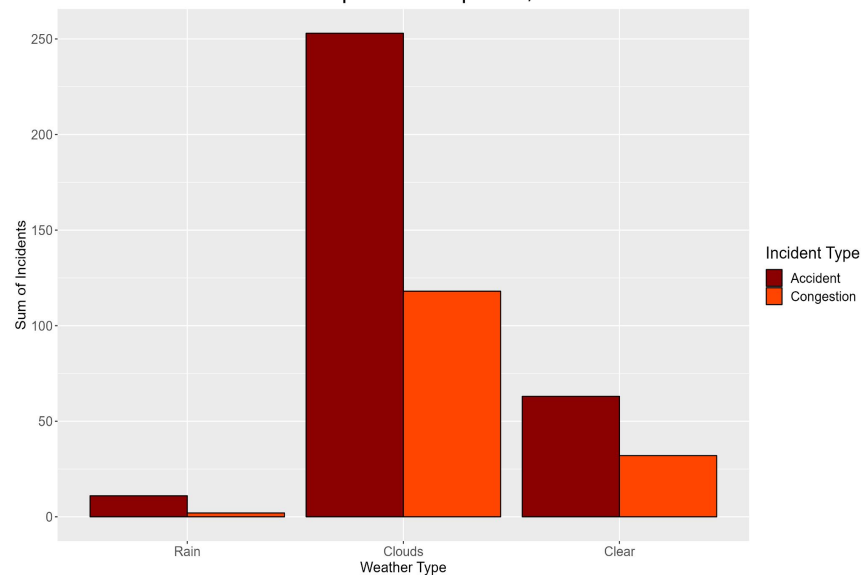


EDA: Sum of Incidents

Seattle Traffic Incidents April 7th to April 21, 2024

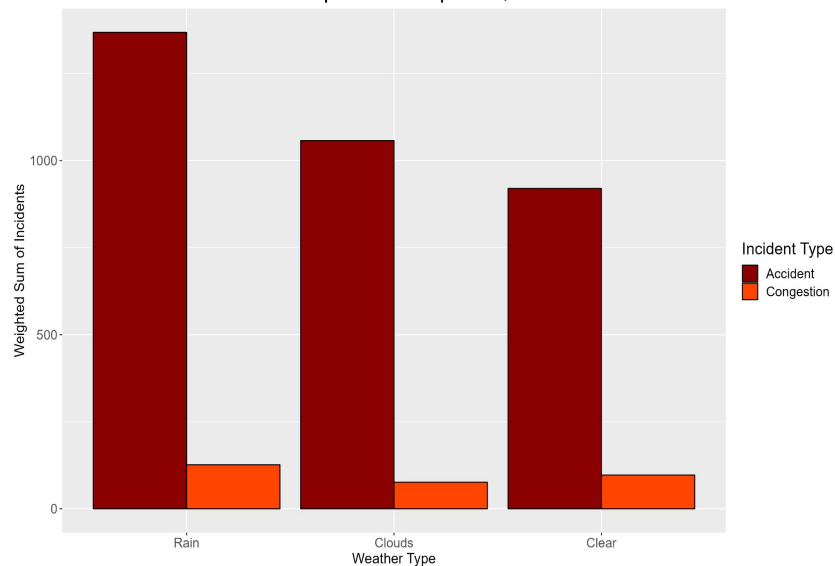


Portland Traffic Incidents April 7th to April 21, 2024

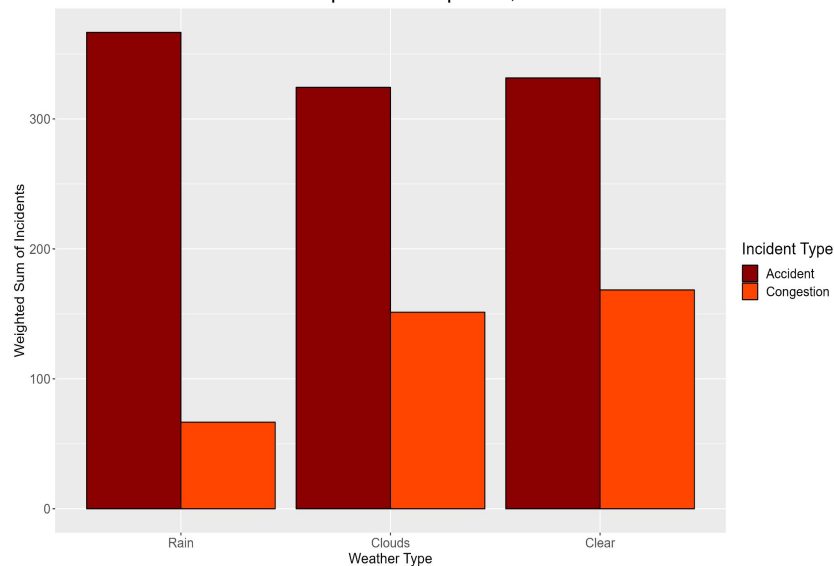


Results: Weighted Analysis

Seattle Traffic Incidents April 7th to April 21, 2024



Portland Traffic Incidents April 7th to April 21, 2024



Conclusions, Limitations, and Further Questions

- It appears that rainy weather does correlate with more accidents
- Two weeks is a small sample size especially when only pulling every 4 hours
- Cloudy was by far the most common weather type
- Timing of pulls doesn't exactly align with time of traffic incidents
- With more data, volume of rainfall could be taken into account
- Looking at cities where rain is less common could be interesting
- Set up more complicated API endpoints