

# Final presentation

By Andrew Arkhipov and Jenny Fotso

# Area of Interest

- Friend did research in flight delays modeled by graphs
- Questions:
  - Do storms delay flights?
    - If so, why, and by how much?
  - How can we prevent storms to help mitigate losses in the airline industry?
    - If so, why?
  - Do storms target airplanes?
    - If so, why?

# Datasets

- Cities: table of all cities in the US
  - Accessed from [simplemaps.com](https://simplemaps.com)
  - Used to augment the storms dataset
- Storms: database of storm events in 2020
  - Accessed from the National Oceanic and Atmospheric Administration
- Delays: database of all flight delays in 2020
  - Accessed from the Bureau of Transportation Statistics

# Cities staging table

|              |         |          |
|--------------|---------|----------|
| city         | STRING  | NULLABLE |
| city_ascii   | STRING  | NULLABLE |
| state_id     | STRING  | NULLABLE |
| state_name   | STRING  | NULLABLE |
| county_fips  | INTEGER | NULLABLE |
| county_name  | STRING  | NULLABLE |
| lat          | FLOAT   | NULLABLE |
| lng          | FLOAT   | NULLABLE |
| population   | INTEGER | NULLABLE |
| density      | INTEGER | NULLABLE |
| source       | STRING  | NULLABLE |
| military     | BOOLEAN | NULLABLE |
| incorporated | BOOLEAN | NULLABLE |
| timezone     | STRING  | NULLABLE |
| ranking      | INTEGER | NULLABLE |
| zips         | STRING  | NULLABLE |
| id           | INTEGER | NULLABLE |

# Delays staging table

|                       |         |          |
|-----------------------|---------|----------|
| YEAR                  | INTEGER | NULLABLE |
| MONTH                 | INTEGER | NULLABLE |
| DAY_OF_MONTH          | INTEGER | NULLABLE |
| FL_DATE               | DATE    | NULLABLE |
| OP_UNIQUE_CARRIER     | STRING  | NULLABLE |
| OP_CARRIER_AIRLINE_ID | INTEGER | NULLABLE |
| TAIL_NUM              | STRING  | NULLABLE |
| OP_CARRIER_FL_NUM     | INTEGER | NULLABLE |
| ORIGIN_AIRPORT_ID     | INTEGER | NULLABLE |
| ORIGIN_AIRPORT_SEQ_ID | INTEGER | NULLABLE |
| ORIGIN_CITY_MARKET_ID | INTEGER | NULLABLE |
| ORIGIN_CITY_NAME      | STRING  | NULLABLE |
| ORIGIN_STATE_ABR      | STRING  | NULLABLE |
| ORIGIN_STATE_NM       | STRING  | NULLABLE |
| DEST_AIRPORT_ID       | INTEGER | NULLABLE |
| DEST                  | STRING  | NULLABLE |
| DEST_CITY_NAME        | STRING  | NULLABLE |
| DEST_STATE_ABR        | STRING  | NULLABLE |

|                     |         |          |
|---------------------|---------|----------|
| DEST_STATE_NM       | STRING  | NULLABLE |
| DEP_TIME            | INTEGER | NULLABLE |
| DEP_DELAY           | FLOAT   | NULLABLE |
| DEP_DELAY_NEW       | FLOAT   | NULLABLE |
| DEP_DEL15           | FLOAT   | NULLABLE |
| DEP_DELAY_GROUP     | INTEGER | NULLABLE |
| WHEELS_OFF          | INTEGER | NULLABLE |
| WHEELS_ON           | INTEGER | NULLABLE |
| TAXI_IN             | FLOAT   | NULLABLE |
| ARR_TIME            | INTEGER | NULLABLE |
| ARR_DELAY           | FLOAT   | NULLABLE |
| ARR_DELAY_NEW       | FLOAT   | NULLABLE |
| ARR_DEL15           | FLOAT   | NULLABLE |
| ARR_DELAY_GROUP     | INTEGER | NULLABLE |
| CANCELLED           | FLOAT   | NULLABLE |
| ACTUAL_ELAPSED_TIME | FLOAT   | NULLABLE |
| AIR_TIME            | FLOAT   | NULLABLE |
| FLIGHTS             | FLOAT   | NULLABLE |
| DISTANCE            | FLOAT   | NULLABLE |
| CARRIER_DELAY       | FLOAT   | NULLABLE |

|                     |         |          |
|---------------------|---------|----------|
| WEATHER_DELAY       | FLOAT   | NULLABLE |
| NAS_DELAY           | FLOAT   | NULLABLE |
| SECURITY_DELAY      | FLOAT   | NULLABLE |
| LATE_AIRCRAFT_DELAY | FLOAT   | NULLABLE |
| FIRST_DEP_TIME      | INTEGER | NULLABLE |
| TOTAL_ADD_GTIME     | FLOAT   | NULLABLE |
| LONGEST_ADD_GTIME   | FLOAT   | NULLABLE |
| string_field_45     | STRING  | NULLABLE |

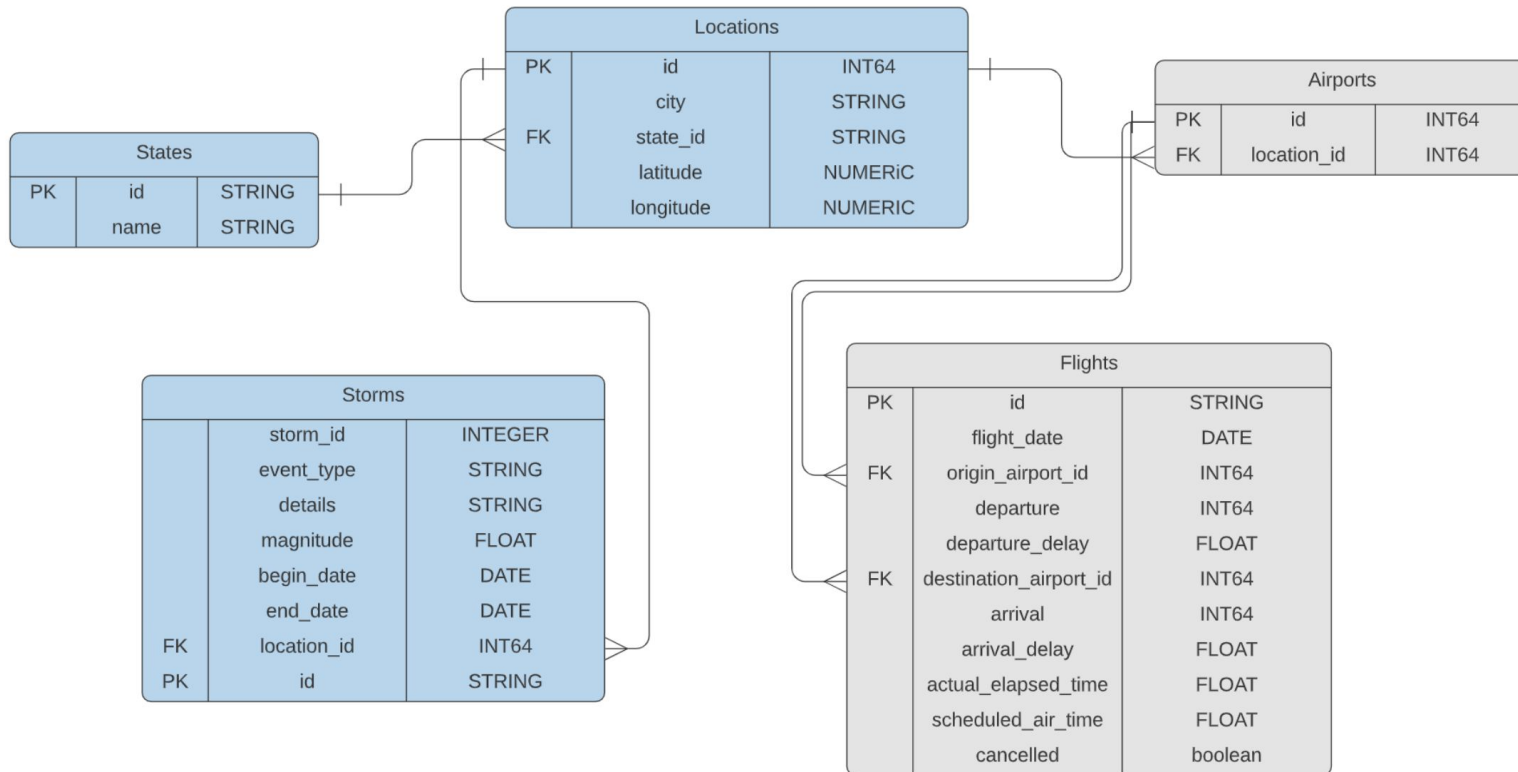
# Storms staging table

|                 |         |          |
|-----------------|---------|----------|
| BEGIN_YEARMONTH | INTEGER | NULLABLE |
| BEGIN_DAY       | INTEGER | NULLABLE |
| BEGIN_TIME      | INTEGER | NULLABLE |
| END_YEARMONTH   | INTEGER | NULLABLE |
| END_DAY         | INTEGER | NULLABLE |
| END_TIME        | INTEGER | NULLABLE |
| EPISODE_ID      | INTEGER | NULLABLE |
| EVENT_ID        | INTEGER | NULLABLE |
| STATE           | STRING  | NULLABLE |
| STATE_FIPS      | INTEGER | NULLABLE |
| YEAR            | INTEGER | NULLABLE |
| MONTH_NAME      | STRING  | NULLABLE |
| EVENT_TYPE      | STRING  | NULLABLE |
| CZ_TYPE         | STRING  | NULLABLE |
| CZ_FIPS         | INTEGER | NULLABLE |
| CZ_NAME         | STRING  | NULLABLE |
| WFO             | STRING  | NULLABLE |
| BEGIN_DATE_TIME | STRING  | NULLABLE |

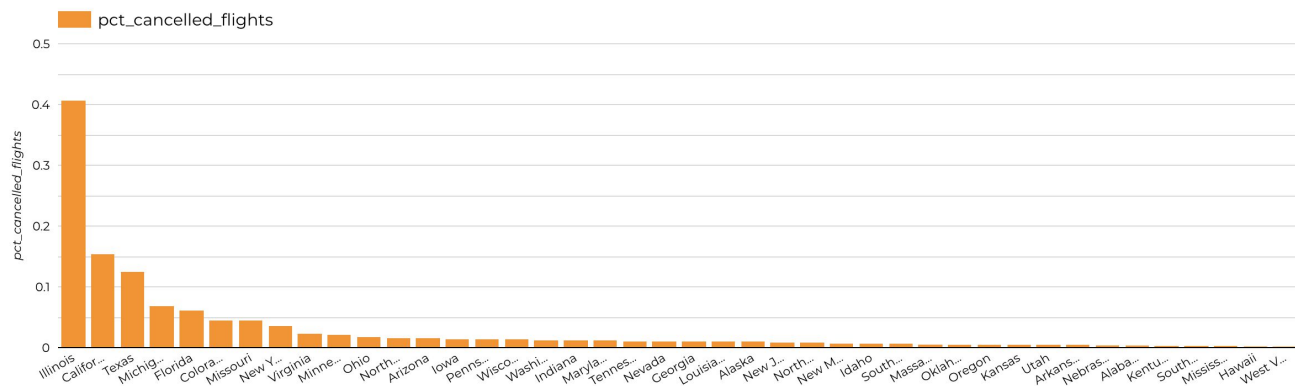
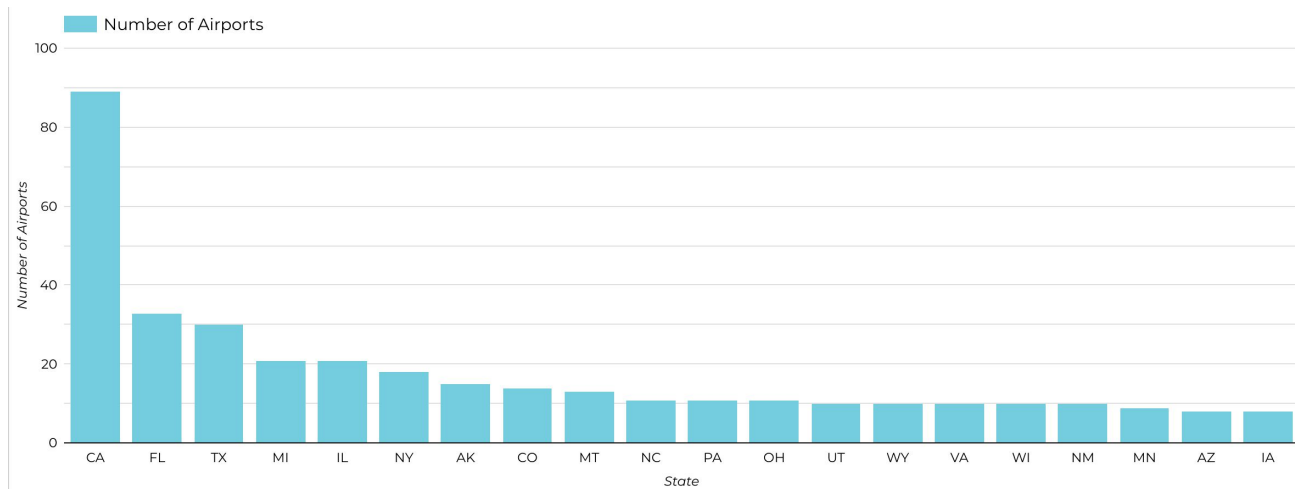
|                    |         |          |
|--------------------|---------|----------|
| CZ_TIMEZONE        | STRING  | NULLABLE |
| END_DATE_TIME      | STRING  | NULLABLE |
| INJURIES_DIRECT    | INTEGER | NULLABLE |
| INJURIES_INDIRECT  | INTEGER | NULLABLE |
| DEATHS_DIRECT      | INTEGER | NULLABLE |
| DEATHS_INDIRECT    | INTEGER | NULLABLE |
| DAMAGE_PROPERTY    | STRING  | NULLABLE |
| DAMAGE_CROPS       | STRING  | NULLABLE |
| SOURCE             | STRING  | NULLABLE |
| MAGNITUDE          | FLOAT   | NULLABLE |
| MAGNITUDE_TYPE     | STRING  | NULLABLE |
| FLOOD_CAUSE        | STRING  | NULLABLE |
| CATEGORY           | STRING  | NULLABLE |
| TOR_F_SCALE        | STRING  | NULLABLE |
| TOR_LENGTH         | FLOAT   | NULLABLE |
| TOR_WIDTH          | INTEGER | NULLABLE |
| TOR_OTHER_WFO      | STRING  | NULLABLE |
| TOR_OTHER_CZ_STATE | STRING  | NULLABLE |
| TOR_OTHER_CZ_FIPS  | INTEGER | NULLABLE |

|                   |         |          |
|-------------------|---------|----------|
| TOR_OTHER_CZ_NAME | STRING  | NULLABLE |
| BEGIN_RANGE       | INTEGER | NULLABLE |
| BEGIN_AZIMUTH     | STRING  | NULLABLE |
| BEGIN_LOCATION    | STRING  | NULLABLE |
| END_RANGE         | INTEGER | NULLABLE |
| END_AZIMUTH       | STRING  | NULLABLE |
| END_LOCATION      | STRING  | NULLABLE |
| BEGIN_LAT         | FLOAT   | NULLABLE |
| BEGIN_LON         | FLOAT   | NULLABLE |
| END_LAT           | FLOAT   | NULLABLE |
| END_LON           | FLOAT   | NULLABLE |
| EPISODE_NARRATIVE | STRING  | NULLABLE |
| EVENT_NARRATIVE   | STRING  | NULLABLE |
| DATA_SOURCE       | STRING  | NULLABLE |

# Modeled tables

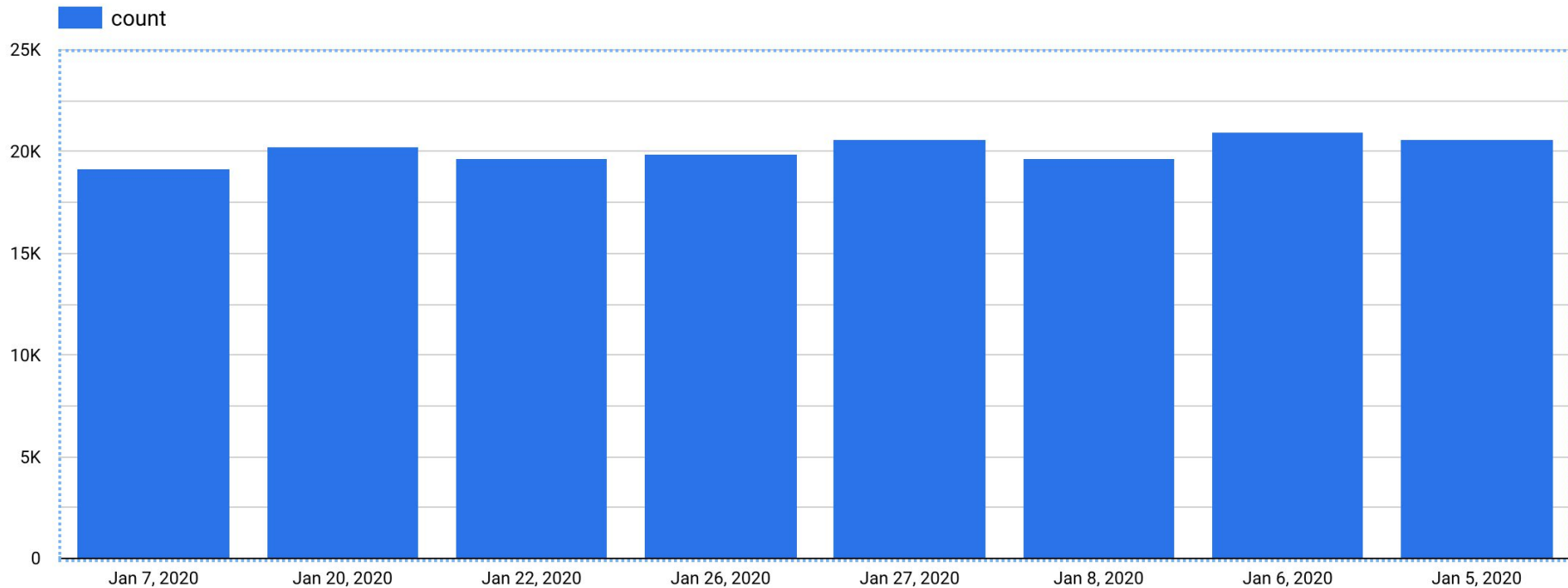


# Number of airports and percentage of cancelled flights by state





## The only days in 2020 with no storms (and respective number of total US flights)



# Reflections

- Results and accomplishments
  - Illinois had a disproportionate number of cancellations, why?
  - Only 8 days between Jan 1 and Aug 31 had no recorded storms in the US
    - Would like to look into doing a hypothesis difference in means test on this
- Challenges
  - Started with a different dataset that didn't have very interesting data, so had to pivot
  - Difficult coming up with meaningful results due to sped up timeline and limited data
  - Doing data analysis in SQL was tough since SQL is very rigid
- Future work
  - Determine by how much exactly storms delay flights