

Are flights going to/coming from the North delayed more?

Databases Group Project WS 23/24

Group U: *Wilson Liu, Cajus Marvin Schneider*

Objective 1: Wilson Liu

Objective 2: Cajus Schneider

Objective 3: Wilson Liu

Objective 4: Cajus Schneider

Concept

Dataset: 2015 Flight Delays and Cancellations ([link](#))

- **Research problem:** We would like to investigate whether flights coming from/going to more northern airports typically have longer delays and to what extent? Our current plan to answer this is by ingesting the data from .csv to SQL files, making the relevant queries on airport latitude, delay time, etc., and graphing the relationship using libraries like Pandas or Seaborn.
- **Motivation:** In Canada there is a common rumor, concerning this research problem. We are curious to see whether this common rumor is based in fact and statistics.
- **Method:** A relational database is ideal for this project as we will need to search for and connect information about two different entities – airports and flights. Also, you have multiple entities with more complex relationships (e.g., 1:M), which would not naturally lend itself to being represented in a spreadsheet or just one .csv file. The airlines and airports can be saved as foreign keys in the *flights* relation, which naturally comes from the dataset.