**ETL Project Report: Airport Weather Delays**  
by  
Kayla St. Germaine, April Lagnevall, Stefano Imperiale

February 2nd, 2021

The primary focus of our project centered on the weather delays for the New York state airports. The availability of data for that selection was the main impetus for continuing research. In order to obtain this information several methods were used under the Extraction step:

* World Weather Online API generation was the first resource needed to begin compiling data.
* We then used World Weather Online to obtain the archived weather data in combination with an Airport CSV obtained by Kaggle, link provided in the bibliography.

Through the use of Pandas via Jupyter Notebook, the data was verified for duplicates, blank fields, and sorted down to a manageable list in order to standardize the results to a more uniform format. On this note, separating the airports into large, medium, and small rankings for the whole state was of particular use. Other points of contention for this step included:

* Latitude and longitude, location
* Type of Airfield, this expanded airports to include facilities such as helipads and closed fields that may not have been indicated at first
* Delay type, some delays were not weather-related
* Search constraints due to the number of pulls from sources that could be made for a day’s data usage cap for API keys

The ‘Load’ step was made possible by using a mixture of PostgresSQL in PG Admin and the SQLAlchemy functions in Jupyter Notebook. The formation of the tables in this way gave us creative control over its structure and layout. By the end of this process, a comprehensive list of 2015 New York state airport delays was created with pertinent information for future clients to view.