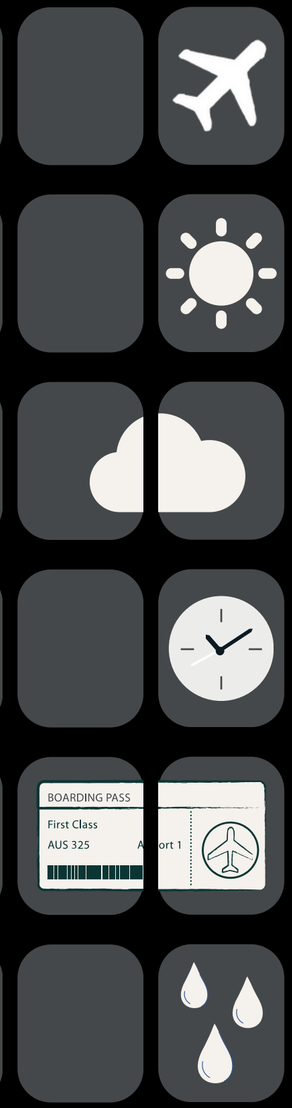
DIFFERENT AIRPORTS AND WEATHER CONDITIONS: HOW FLIGHT DELAYS CHANGE?



group: Adobati, Ghiselli Ricci, Mariani academic year: 2023-2024 course: data management

Could weather conditions influence flight departure delays?

DATA ACQUISITION

collecting flights data and weather information

DATA CLEANING

considering only the non-superfluous features

DATA INTEGRATION

merging and adding information to create a unified dataset

DATA STORAGE

storing data in a document-based database called MongoDB

DATA QUALITY

measuring data quality by evaluating completeness, consistency, currency

ANALYSIS AND QUERIES

general study providing tools for further analysis

DATA ACQUISITION

- **What?**

  We chose 7 airports based on different geographical and climate areas:

  Artic climate - *Reykjavik Airport*
  High mountain climate - *El Dorado International Airport*
  Mediterranean climate - *Athens International Airport*
  Subtropical climate - *Miami International Airport*
  Temperate continental climate - *Milan Malpensa Airport*
  Temperate oceanic climate - *Narita International Airport*
  Tropical climate: *Ninoy Aquino International Airport*

- **When?**

  Starting from 7th of December 2023 to 14th of February 2024, the implemented code runs every 30 minutes

- **How?**

  Different methods: API for climate data, scraping for airports and flights data and file download for IATA codes acronyms

DATA ACQUISITION

API (Application Programming Interface)

We found a public API called Open-Meteo that gives detailed weather conditions at precise point using geospatial coordinates and time

The endpoint is *https://archiveapi.open-meteo.com/v1/archive* and needed parameters are date, time, longitude, latitude and the list of measures wanted

Considered measures:

- precipitation [mm]
- cloud-cover [%]
- wind speed at 10m [km/h]
- wind speed at 100m [km/h]

The API provides data hourly, so we did a weighted mean to approximate data at a specific minute

DATA ACQUISITION

Scraping

Since there are no public APIs for flights data, we examined the distinct requests from the browser of each airport page

Types of response are JSON or HTML (for these documents we used BeautifulSoap library)

The features of our interest are:

- flight-number
- flight-status
- scheduled-departure
- actual-departure
- airport-departure
- airport-arrival

To link the airports with their corresponding coordinates, we did scraping on a website which returns latitude and longitude in JSON format

DATA ACQUISITION

File Download

Both the name of the flight departure and arrival airports are encoded into an acronym using a standard called IATA

The file dowloaded from Internet contains:

- acronyms associated with the corresponding airport
- the city where the airport is located
- the state where the airport is located

DATA CLEANING

Some data can be semantically superfuous for our research question even though they are considered correct in terms of form and syntax

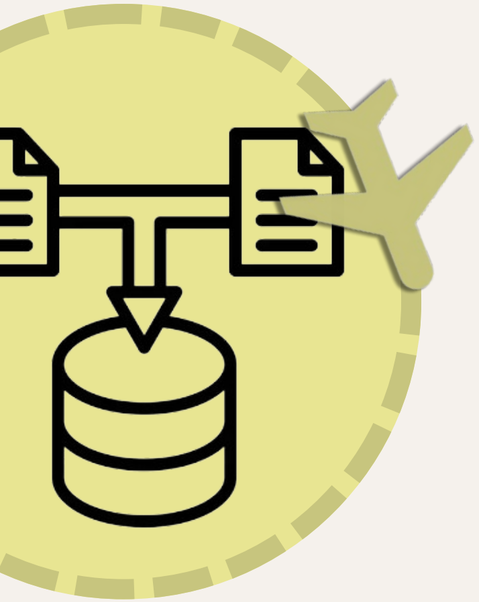
For this reason, there are many features that we can exclude from our dataset

We kept only attributes that were useful for our reasearch question. They are structured as follows:

- number - *EC03911*
- status - *DEPARTED* or *CANCELLED*
- scheduledDep - *2023-12-07T06:00:00.000+00:00*
- actualDep - *2023-12-07T06:19:00.000+00:00*
- airportDep - *MXP*
- airportArr - *Luxembourg Airport*

The IATA file contained some formatting errors and a missing name

DATA INTEGRATION

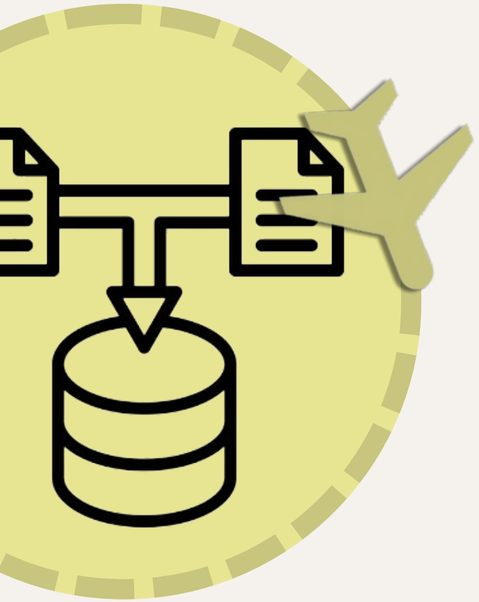


Integration

After finding out where each value was located in each scraping response, we merged data to uniform the output of the process in a single dataset

We also associated all flights with the corresponding weather conditions at departure time, using geospatial coordinates of each airport

DATA INTEGRATION



Enrichment

To improve the comprehension, we substituted the plain airport name with acronym provided by IATA standard

Another thing we have done is to create a uniform way to represent *number*, by taking the airline name and the flight number given by the airport

DATA STORAGE

We decided to implement a NoSQL document-based database, MongoDB.
The type of implementation we rely on, is the remote hosting service offered by MongoDB platform, called Atlas.
This service provides a free to use database with some limitations, like storage limit or bandwidth limit.

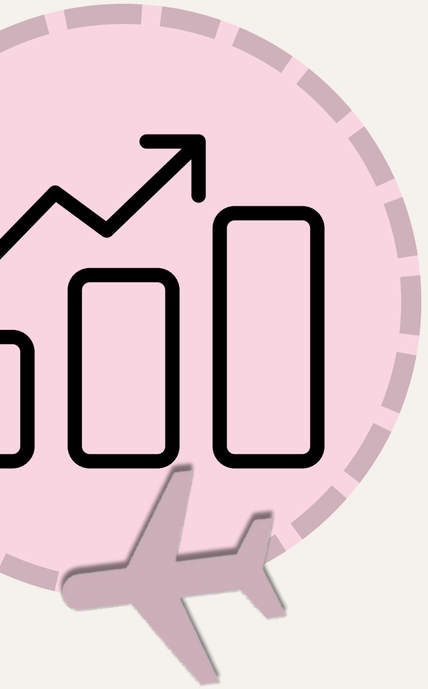
We created two collections that store flights and IATA information, *collflights* and *colliata.*

DATA QUALITY

We selected different quality dimensions, completeness, consistency and currency, which are the most appropriate for our system analysis.

- Completeness, evaluated on the flight actual departure time, is **98,652%**
- Consistency, evaluated on the arrival airport name is **90,260%**
- Currency is **30 minutes** for flights data and **2 days** for weather conditions data

ANALYSIS&QUERIES

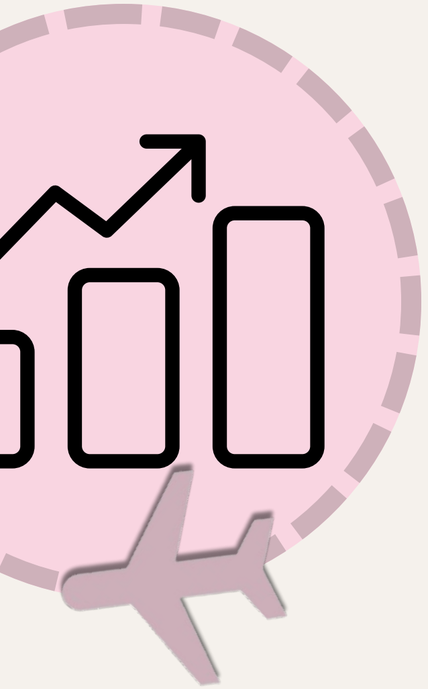


Analysis

To answer our research question, we did a descriptive analysis to understand which one of the weather condition measures mostly influence flight delays.
The results are illustrated in the correlation matrix on the right

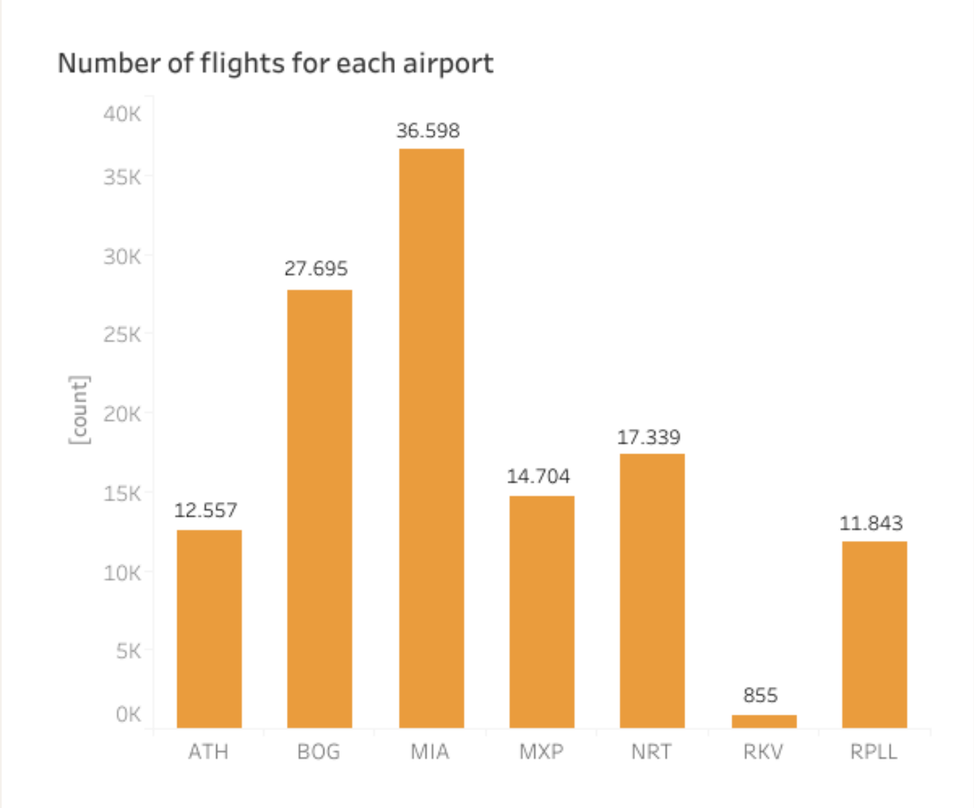
	Precipitation	Cloud	Wind 10m	Wind 100m
ATH	0,015	0,003	0,043	0,038
BOG	0,004	-0,012	-0,002	-0,023
MIA	0,006	0,025	0,001	0,002
MXP	0,046	0,031	-0,026	-0,010
NRT	0,007	-0,013	0,027	0,006
RKV	0,126	0,071	0,008	0,066
RPLL	-0,001	0,011	-0,012	-0,007

ANALYSIS&QUERIES

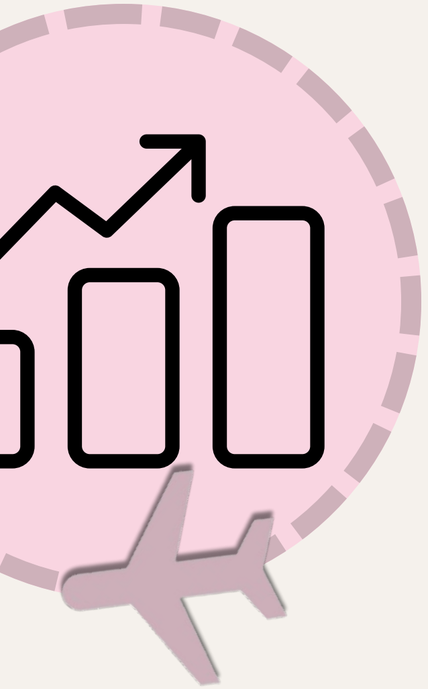


Query 1

Returns the number of flights for each airport

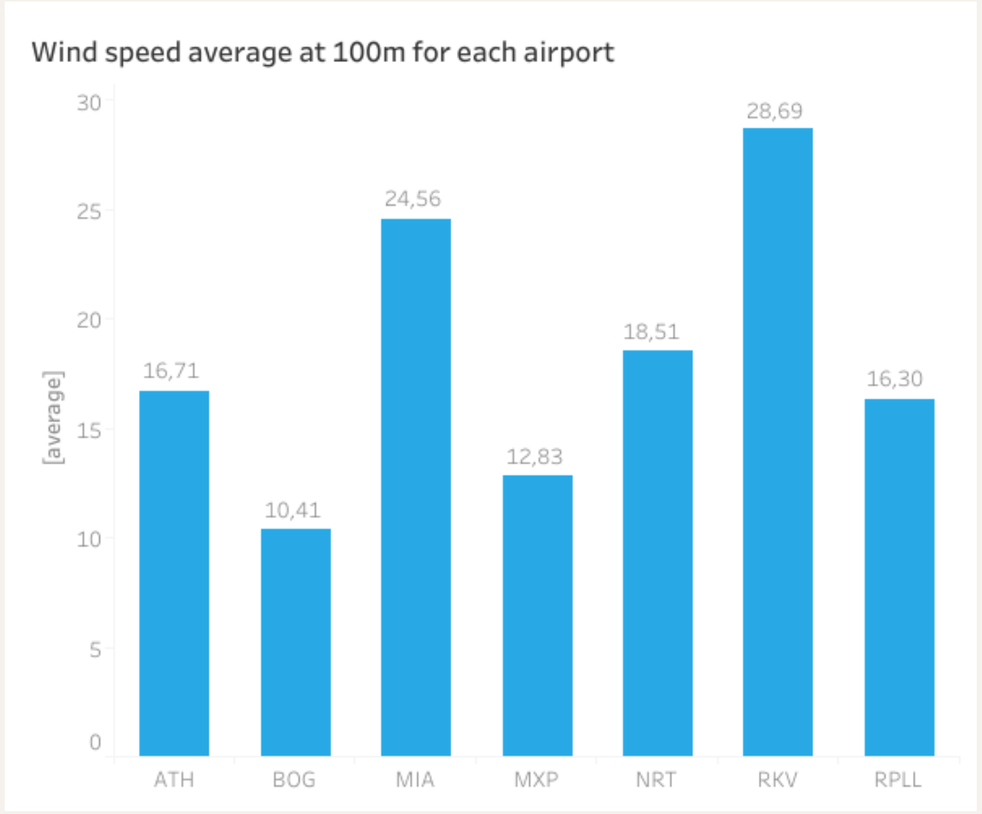


ANALYSIS&QUERIES

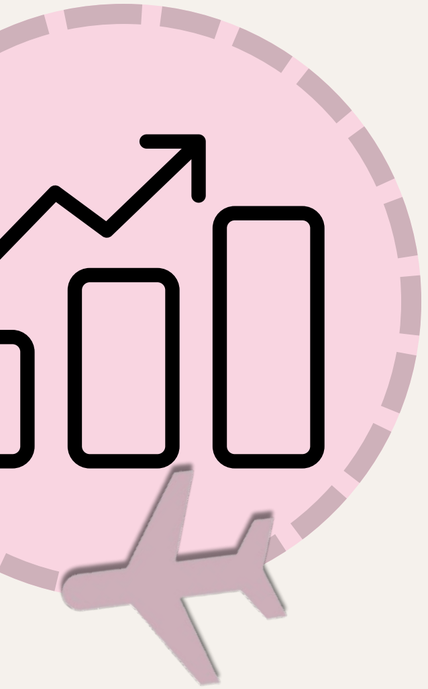


Query 2

Returns the average speed of wind at 100 metres for each airport

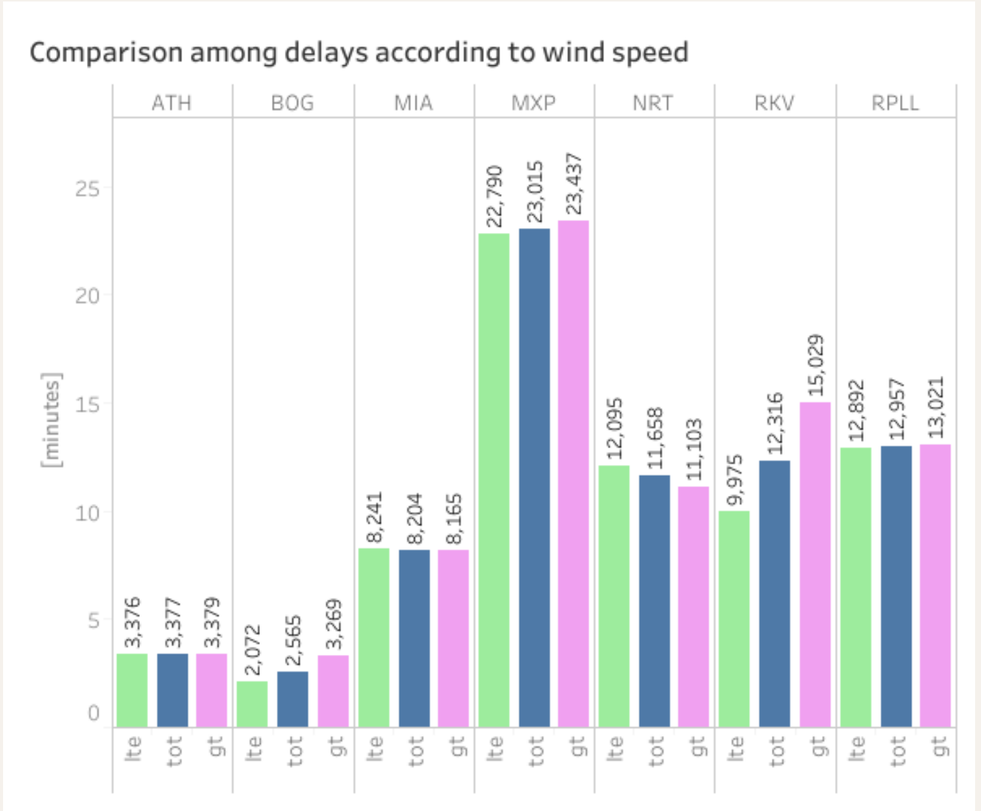


ANALYSIS&QUERIES

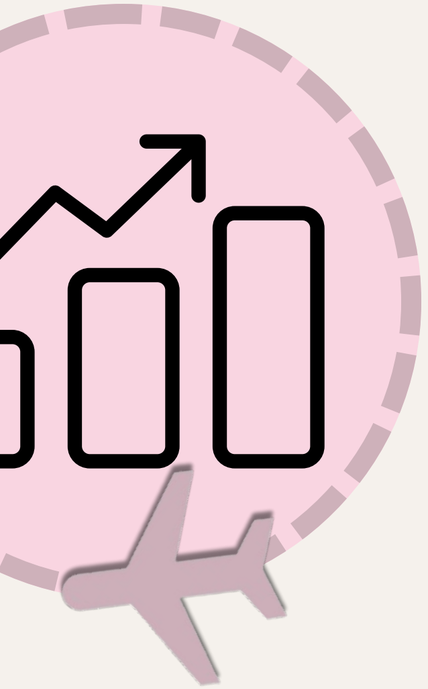


Query 3

Shows the comparison among delays according to wind speed at 100 metres for each airport

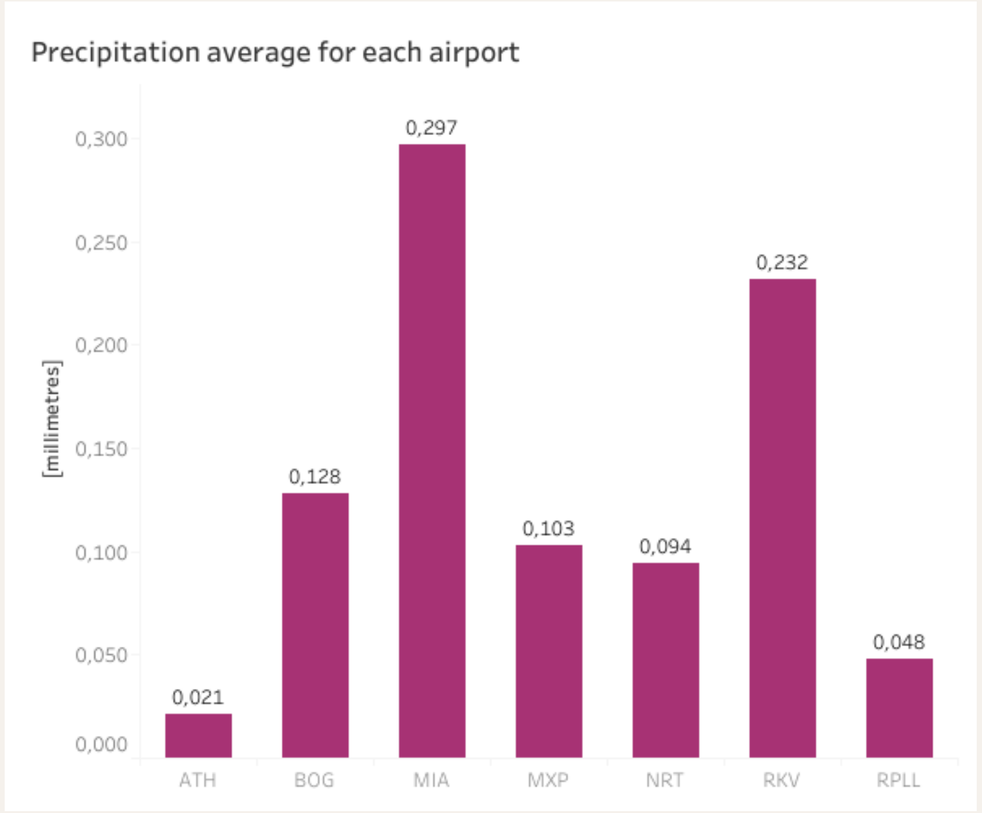


ANALYSIS&QUERIES

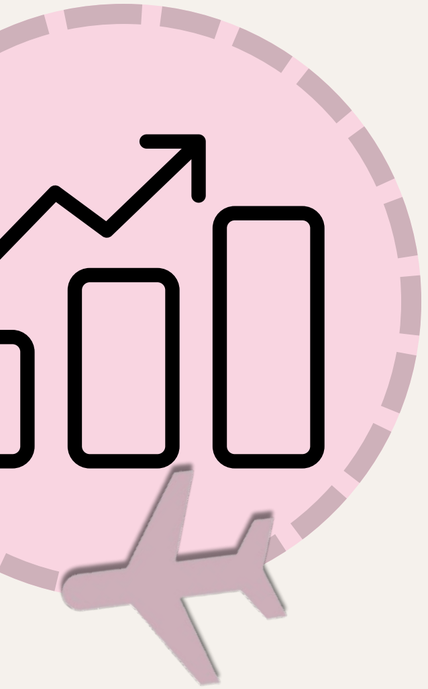


Query 4

Returns the average of precipitation for each airport

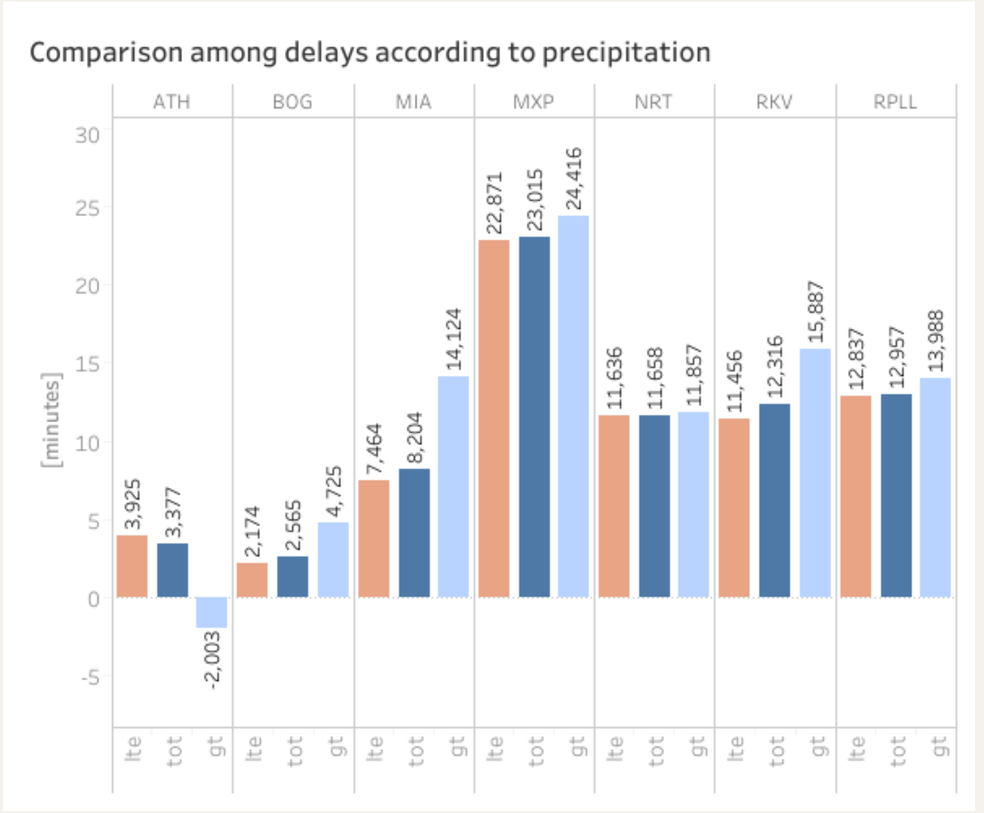


ANALYSIS&QUERIES

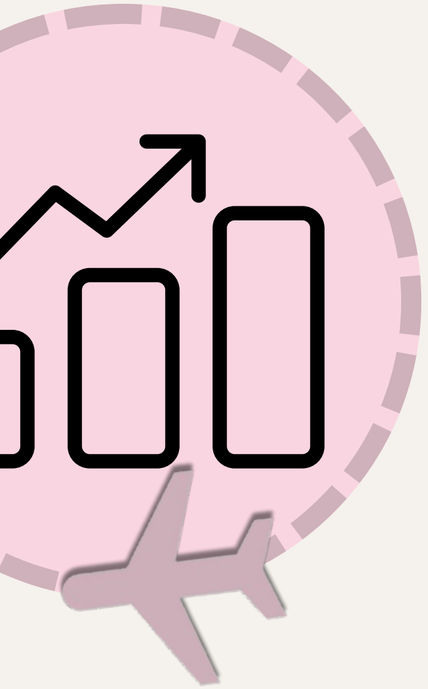


Query 5

Shows the comparison among delays according to precipitation for each airport



ANALYSIS&QUERIES



Percentage increase

This is a recap on how delays change in each airport according to the previous queries

	Wind	Precipitation
ATH	+0,089%	-151,032%
BOG	+55,77%	+117,341%
MIA	-0.922%	+89,228%
MXP	+2,839%	+6,755%
NRT	-8,202%	+1,899%
RKV	+50,667%	+38,678%
RPLL	+1,001%	+8,966%

THANKS FOR THE ATTENTION!