Snowflake Workshop

Elisabetta Parozzi & Luca Pescatore

Argusa

4/12/2025

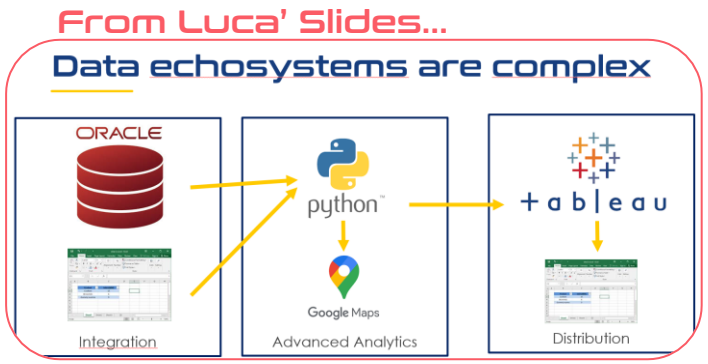
Getting Started

WHY SNOWFLAKE?

- **Scalability & Performance** – Handles large datasets efficiently with automatic scaling and parallel processing.
- **Separation of Storage and Compute** – Enables flexible resource management and cost optimization.
- **Cloud-Native Platform** – Accessible from anywhere, with built-in security, data sharing, and integration features.
- **SQL-Based & User-Friendly** – Leverages standard SQL syntax while supporting advanced analytics and geospatial operations.

Getting Started

WORKSHOP OBJECTIVE



- Learn how to analyze spatial data in Snowflake.
- Work with real Swiss data (students and schools).
- Understand how to join datasets, compute school distances, and derive insights about school accessibility.
- Gain hands-on experience using Snowflake SQL functions for data analysis and transformation.

Steps Overview

1

Load and Source data:
- Download repository
- Load the csv file
- Source the database

2

Filter and Organize:
- Separate residential buildings from Schools
- Combine students and residentials
- Join tables for complete information

3

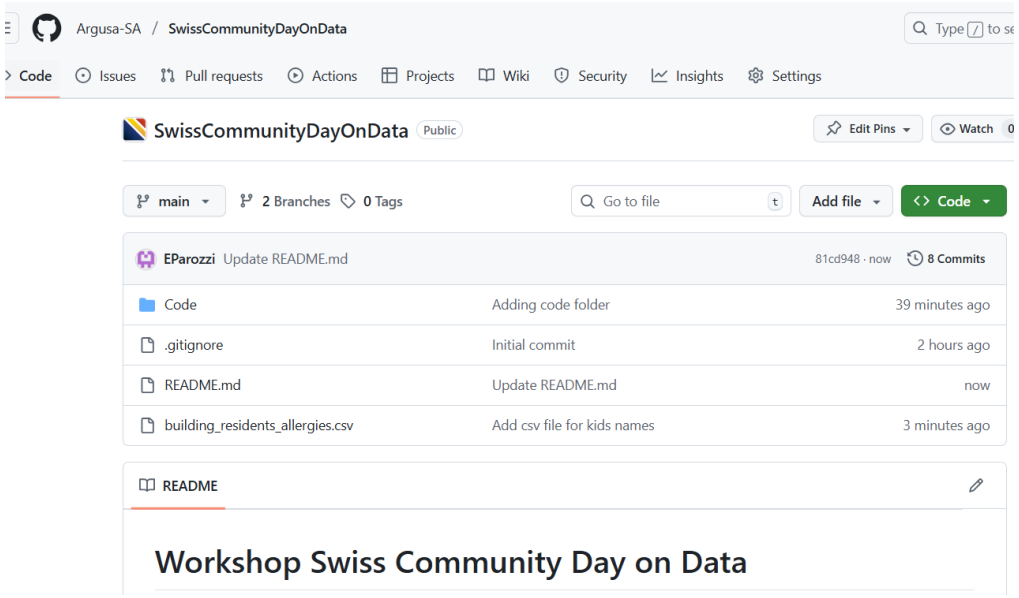
Calculate Distance:
- We use the tables and views created to calculate the distance between each residential building
- Final dataset

4

Chat bot App:
We employ CORTEX to inquiry data with a natural language!

Useful links:

LINK WHERE TO SOURCE THE CODE AND THE CSV FILE.

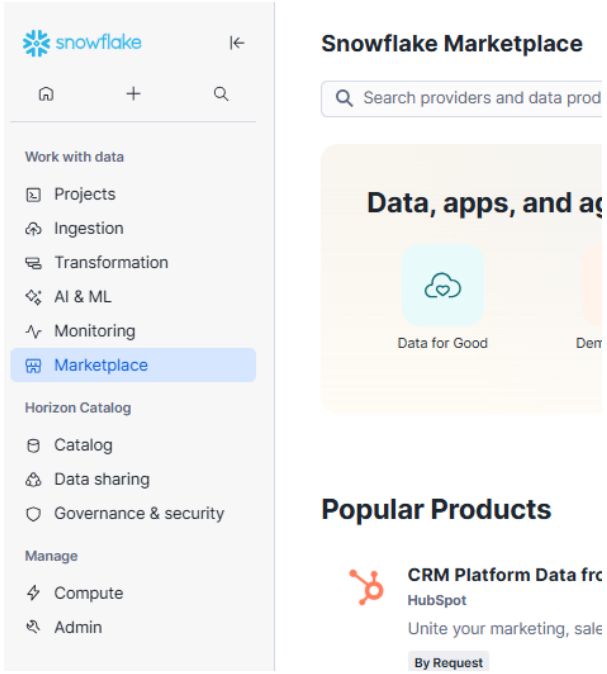


- https://github.com/Argusa-SA/SwissCommunityDayOnData

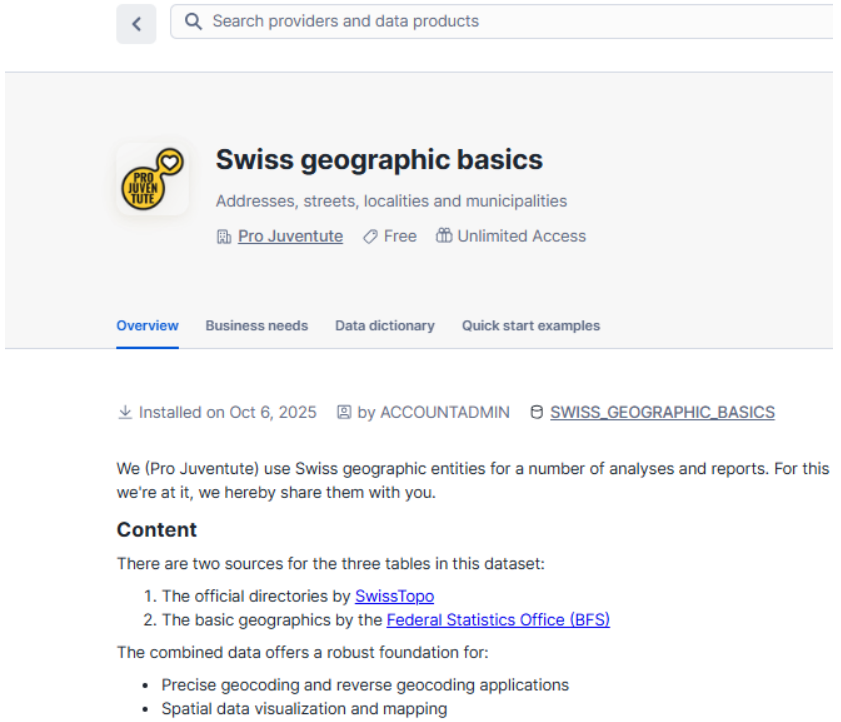
- Let's now open Snowflake!
- **USERNAME: SCDD**
- **Password: Swisscommunitydaydata2025**
- **https://cvhawsm-uj05793.snowflakecomputing.com**

Step 1. Source Database

GO TO THE MARKETPLACE



DOWNLOAD THE FREE DATABASE



Step 1. Add CSV file

LOADING AN EXTERNAL FILE TO THE ANALYSIS

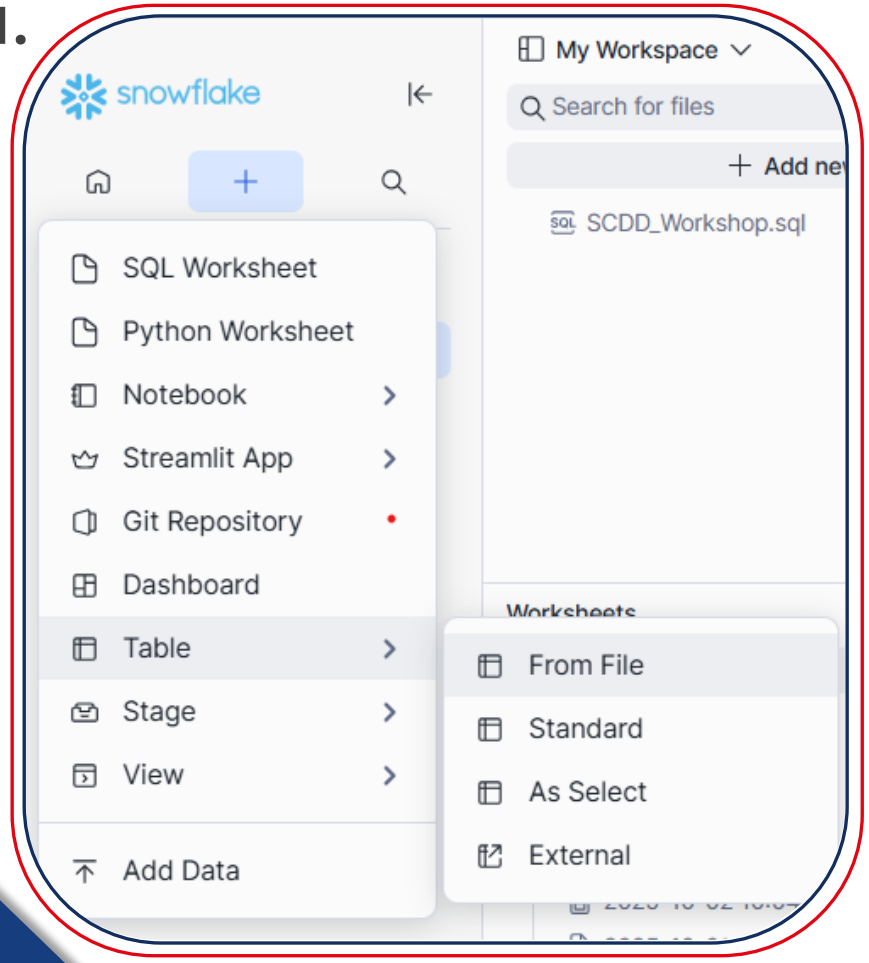
- **1. Upload & Load with the “+” Button**
  - From the Snowflake Web UI, navigate to your database → schema → table.
  - Click the “+” icon → Load Data.
  - Select your local CSV file, define the file format (e.g., comma-delimited, header rows, encoding).
  - Snowflake automatically uploads and loads the data directly into the selected table.

- **2. Load via Stage and COPY INTO Command:**
  - Create or use an existing **internal stage** (@mystage) or **external stage** (e.g., AWS S3).
  - Upload your CSV file to the stage using “COPY INTO” SQL command.

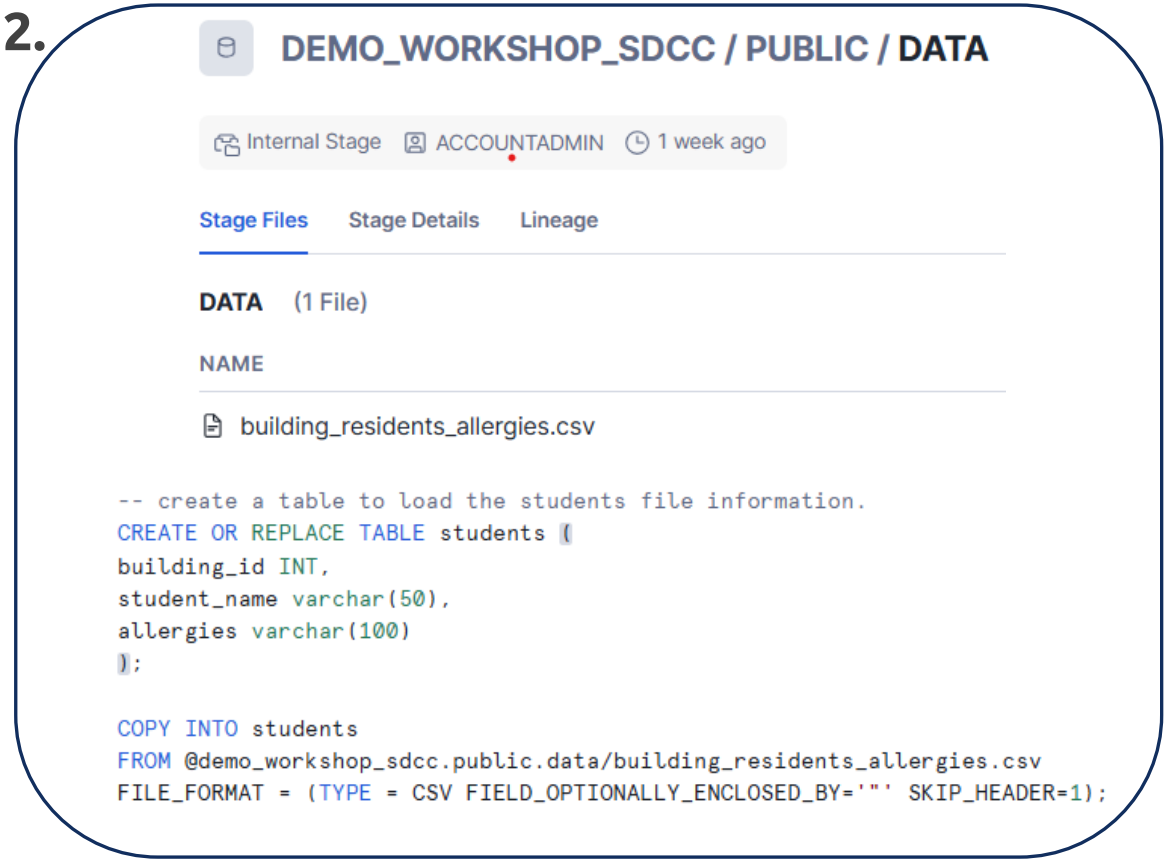
Step 1. Add CSV file

LOADING AN EXTERNAL FILE TO THE ANALYSIS

1.



2.



Step 2. Filter and Organize Data

STRUCTURE AND PREPARE THE RAW DATA FOR ANALYSIS BY ISOLATING RELEVANT SUBSETS (E.G., RESIDENTIAL BUILDINGS, SCHOOLS, STUDENTS).

- **1. Filter the Data by Canton**
  - Extract only the records for Canton Zug (ZG) from the Swiss geographic dataset.
  - Create working tables such as buildings_zg for further processing.

```
SELECT *
FROM swiss_geographic_basics.buildings
WHERE canton = 'ZG';
```

- **2. Identify and Separate Key Building Types**
  - Residential buildings → filter by building_category = 'residential'.
  - Schools → select where building_name ILIKE '%Schul%'.
  - Store them in new tables:
    - residences_zg
    - schools_zg

```
CREATE OR REPLACE TABLE schools_zg AS
SELECT *
FROM buildings_with_locality_zg
WHERE building_name ILIKE '%Schul%';
```

Step 2. Filter and Organize Data

- **3. Enrich Buildings with Municipality Information**
  - Join buildings streets to include geographic context: Municipality
  - Create a clean, enriched table “buildings_with_municipality_zg”.

```
FROM buildings_zg b
JOIN streets_zg s
  ON b.street_id = s.street_id;
```

- **4. Add Student Information**
  - Join residences_zg with the students table imported from the CSV file
  - Result: buildings_zg_students — a dataset linking each student to their residence and allergies.

```
JOIN students s
ON r.building_id = s.building_id;
```

Step 3. What are we looking for?

DETERMINE THE DISTANCE BETWEEN EACH STUDENT'S RESIDENCE AND THE NEAREST SCHOOL IN CANTON ZUG, USING SNOWFLAKE'S GEOSPATIAL FUNCTIONS.

- **1. Compute Geospatial Distances**
  - Use the ST_DISTANCE() function to calculate the distance (in meters) between each residence and nearby schools.
  - Filter for the closest school using a ranking function:

```
QUALIFY ROW_NUMBER() OVER (
    PARTITION BY r.building_id
    ORDER BY ST_DISTANCE(r.POINT_WGS84, s.POINT_WGS84)
) = 1;
```

- **2. Create the Distance Table**
  - The result is stored in residence_school_mart_zg, containing:
    - Residence details (address, municipality, student, allergies)
    - Closest school name and ID
    - Calculated distance_meters

Step 3. What are we looking for?

RESULTS

```
SELECT MUNICIPALITY,
       AVG(distance_meters) AS avg_distance,
       MEDIAN(distance_meters) AS median_distance
FROM residence_school_mart_zg
GROUP BY MUNICIPALITY;
```

- **3. Analyze Accessibility Insights**
  - Aggregate results by municipality.
  - Identify areas where **students live farther from schools** → potential insight for **urban planning or accessibility studies**.

Results (just now)

Table | Chart | 92 rows | 923ms

	RESIDENCE_ID	STREET	STUDENT_NAME	ALLERGIES	MUNICIPALITY	ZIP_LOCALITY	BUILDING_CANTON	SCHOOL_ID
	310291 19207...	Sch... 3.3% Bah... 2.2% +80 more	Alessandr... 1.1% Alessandr... 1.1% +90 more	Fish 9.8% None 9.8% +18 more	Zug 23.9% Baar 20.7% +9 more	6340 Baar 29.3% 6300 Zug 19.6% +14 more	ZG 100.0%	9033... 1920...
1	190948129	Carmelweg	Oskar Kaiser	Peanuts, Tree nuts	Neuheim	6313 Menzingen	ZG	9033453
2	3155347	Am Rainbach	Heinrich Sommer	Eggs	Neuheim	6340 Baar	ZG	192032281
3	312675	Bahnhofstrasse	Helga Schwab	Tree nuts, Eggs	Zug	6330 Cham	ZG	192032281
4	191736807	Mühlegasse	Felix Eder	Gluten	Baar	6340 Baar	ZG	192032281
5	321011	Loretohöhe	Gabriel Petit	Gluten	Baar	6300 Zug	ZG	192032281
6	310606	Rebmattli	Alessandro Colombo	Sesame	Zug	6340 Baar	ZG	192032281
7	312167	Moosmattstrass	Juliette Boyer	Shellfish	Cham	6330 Cham	ZG	192032281

Final Step: Streamlit

LET'S LOOK AT THE STREAMLIT BUILT CHAT-BOT

- **Objective:**
  - Build an interactive Streamlit application connected to Snowflake.
  - Use Cortex Search to query student information and allergy data in natural language.
- **Key Features:**
  - Conversational Interface: Ask questions such as *"Which students in Zug are allergic to peanuts?"*
  - Real-Time Data Access: Retrieves results directly from Snowflake tables.
  - Multilingual Output: Displays translated allergy information (EN, FR, DE, IT).
  - User-Friendly Visualization: Explore schools, residences, and allergy data interactively.

Stats and Feedback

Take a moment to quickly answer to our questions. Your answers will help us gather statistics about today's participants, understand the background and experience levels in the room, and improve future Snowflake workshops.

We appreciate your contribution!

THANK YOU!

Argusa