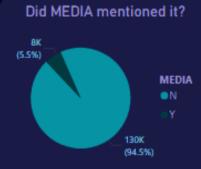


UFO Reports Analysis (1950-2025)



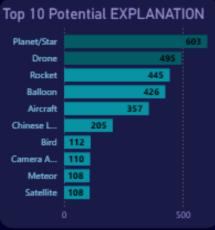
138K

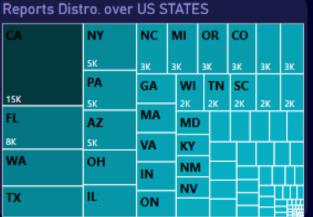




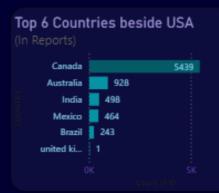














(8 oct 2025)
Presenter\ Maher Mahmoud

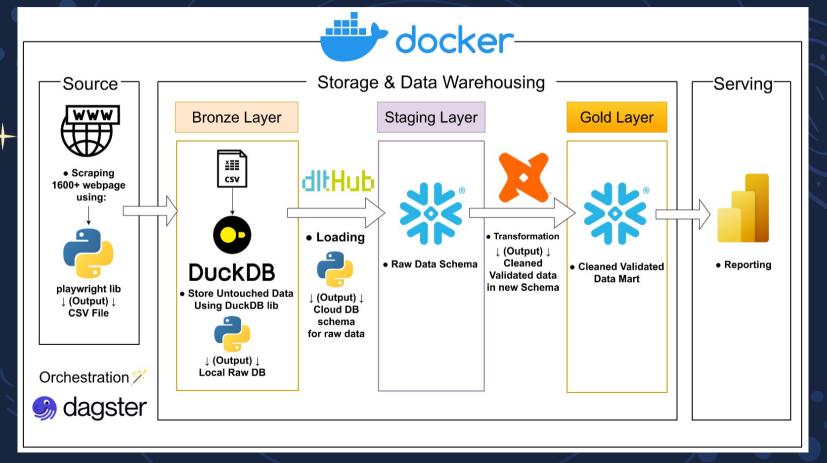




Affected Users

- UFO Hunters
- Paranormal Investigators

Planning phase: Workflow chart



TECHNOLOGY MATRIX ((Tech Stack))



Orchestration

Category

Dagster

Docker

Defining, executing, and visualizing the end-to-end data pipeline.

Purpose

Creating a consistent and portable environment for all scripts and dependencies.

Data Ingestion (Extraction)

Python (Playwright)

Technology / Service

Scraping raw UFO sighting data from the NUFORC public website.

Intermediate Storage

Containerization

DuckDB

Staging the raw scraped data in a local, file-based

analytical database. Loading data from the intermediate DuckDB

Data Ingestion (Loading) Data Warehouse

Load Tool) Snowflake

Python (dlt) (Data

Storing raw (Bronze) and transformed (Silver/Gold) data for analysis.

database into Snowflake.

DBT

Power BI

Cleaning, standardizing, and modeling the data directly within Snowflake.

source code.

Building the final dashboards and reports for data analysis.

Programming Language

Data Transformation

BI & Data Modeling

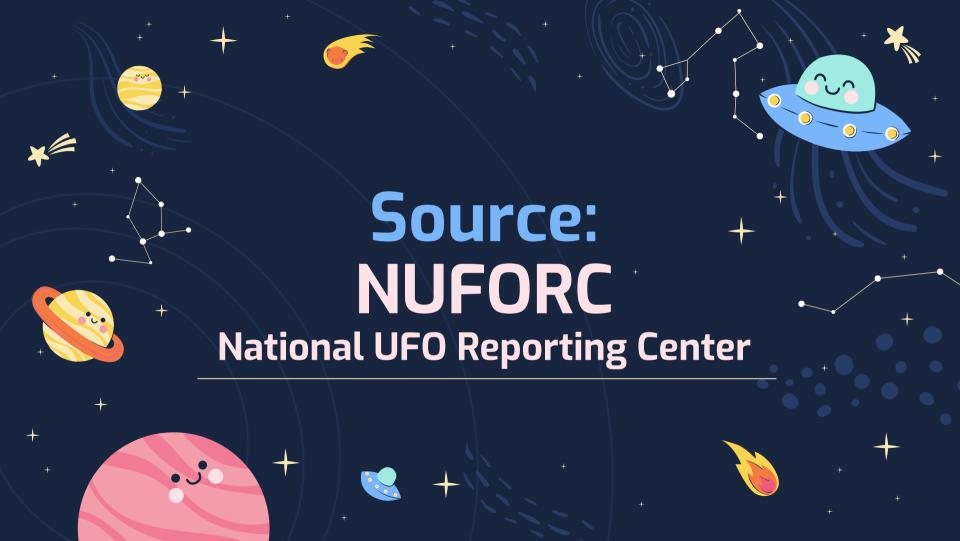
Version Control

Python

Git & GitHub

The core language used for scripting the scraping, orchestration, and data loading scripts. Managing and tracking changes to the project's

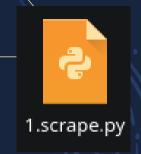








Python Script using Playwright Libarary

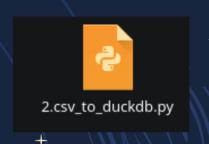








Python Script using DuckDB Libarary





↓Output ↓ Database FILE for local storage







Python Script using DLT (data load tool) Libarary





↓Output ↓ stored Data into cloud Storage





* DBT Scripts

↓ Output↓ Cleaned & Validated Data in new schema

```
tree .
 -- README.md
    dbt packages
   dbt project.yml
  - loas
     -- dbt.loa
 -- models
     -- schema.yml
     -- ufo_reports.sgl <mark>-- Transformation</mark>
 -- profiles.vml
 -- target
     -- compiled
         -- ufo dbt
              -- models
                  -- example
                      -- schema.yml
                          -- not null ufo reports id.sql
                          -- not null ufo reports occurred sql
                          -- unique ufo reports id.sal
                      -- ufo reports.sal
     -- graph.gpickle
     -- graph_summary.json
     -- manifest.json
        partial_parse.msgpack
     -- perf_info.json
     -- run
         -- ufo dbt
             -- models
                 -- example
                      -- schema.yml
                          -- not_null_ufo_reports_id.sql
                          -- not_null_ufo_reports_occurred.sql
                          -- unique ufo reports id.sql
                      -- ufo reports.sal
     -- run results.json
     -- semantic_manifest.json
15 directories, 21 files
```

Transformation process





	n.	⁺ ACT	Reason
	01	Remove "OPEN"	99% of the column is 1 value
+	03	Remove "link"	Unused column in analysis
	04	rm. nulls "occurred"	238 null records removed
	05	rm. Dublicaties	41872 Dupli record removed
•	06	Sort "occurred" and Add Serrogate key	Adding uniqueness identifier to records

Transformation process





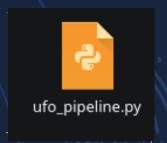
	n.	⁺ ACT	⁺ Reason
	07	Replace null values In "media"	Replace nulls values to "N"
+	08	Replace null values in "explaination"	Replace nulls values to "N/A"
	09	Making "id" a PK	Make id the Primary key
+	10	Normalize values In "explaination"	Cleaning to 91 explaination typo value

		case
	EXPLANATION	when upper(explanation) like '%AIRCRAFT%' then 'Aircraft'
	Definition.	when upper(explanation) like '%BALLOON%' or upper(explanation) like '%BALLOONS%' then 'Balloon'
33	Reflection	when upper(explanation) like '%BAT%' then 'Bat'
34	Contrail	when upper(explanation) like '%BIRD%' or upper(explanation) like '%BIRDS%' then 'Bird'
14	Contrail	when upper(explanation) like '%BLIMP%' then 'Blimp'
35	ISS	when upper(explanation) like '%BOAT%' or upper(explanation) like '%BOATS%' then 'Boat'
	155	when upper(explanation) like '%BUILDING%' then 'Building'
36	Birds?	when upper(explanation) like '%CLOUD%' then 'Cloud'
		when upper(explanation) like '%COMET%' then 'Comet'
37	Truck	when upper(explanation) like '%CONTRAIL%' then 'Contrail'
		when upper(explanation) like '%CRATER%' then 'Crater'
38	Bird	when upper(explanation) like '%DEBRIS%' then 'Debris'
		when upper(explanation) like '%DRONE%' or upper(explanation) like '%DRONES%' then 'Drone'
39	Contrail?	when upper(explanation) like '%FLARE%' or upper(explanation) like '%FLARES%' then 'Flare'
	/	when upper(explanation) like '%FIREWORK%' or upper(explanation) like '%FIREWORKS%' then 'Fireworks'
40	Laser?	when upper(explanation) like '%HEADLIGHT%' or upper(explanation) like '%HEADLIGHTS%' then 'Headlight
41	Ball Lightning?	when upper(explanation) like '%HOAX%' then 'Hoax'
41	Ball Lightning?	when upper(explanation) like '%INSECT%' or upper(explanation) like '%INSECTS%' then 'Insect'
42	Starlink-Racetrack	when upper(explanation) like '%ISS%' then 'ISS'
	Jean III I Jacob de N	when upper(explanation) like '%KITE%' then 'Kite'
43	birds?	when upper(explanation) like '%LANTERN%' or upper(explanation) like '%CHINESE LANTERN%' or upper(exp
		when upper(explanation) like '%LASER%' then 'Laser'
44	Dream?	when upper(explanation) like '%LIGHTNING%' then 'Lightning'
	—————————————————————————————————————	when upper(explanation) like '%METEOR%' then 'Meteor'
45	Aircraft?	when upper(explanation) like '%MOON%' then 'Moon'
	\	when upper(explanation) like '%PLANET%' or upper(explanation) like '%STAR%' then 'Planet/Star'
46	Balloon	when upper(explanation) like '%ROCKET%' then 'Rocket'
	Ralloon?	when upper(explanation) like '%SATELLITE%' or upper(explanation) like '%SATELLITES%' then 'Satellite
47	Balloon? 2	when upper(explanation) like '%SEARCHLIGHT%' then 'Searchlight'
48	Drone?	when upper(explanation) like '%SMOKE%' then 'Smoke'
40	Droner	when upper(explanation) like '%SPACE JUNK%' then 'Space Junk'
49 (Chinese Lantern?	when upper(explanation) like '%STARLINK%' then 'Starlink'
	Chinese Editerni	when upper(explanation) like '%SUNDOG%' then 'Sundog'
50	Starlink (Racetrack)?	when upper(explanation) like '%AURORA%' then 'Aurora'
	- I I I I I I I I I I I I I I I I I I I	when upper(explanation) like '%BALL LIGHTNING%' then 'Ball Lightning'





Python Script using Dagster Libarary



Jobs automation in order ↓
1- Scrapping script
2- Local Storing Script
3- Loading into Cloud script
4- DBT transformation script





github repo clone

github.com/maher-dataconsult /nuforc_pipeline







Docker

1- cd nuforc_pipeline 2- build image 3- run container





Power Query⁺ transformation



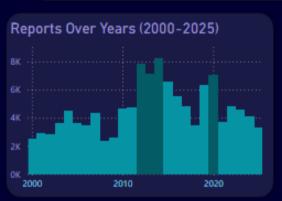


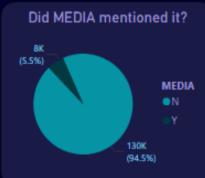
	n.	⁺ ACT	⁺ Reason
	01	Adding "year" column	Extract only Year from "occurred"
+	02	Transform "explaination" Into new column	Replace N/A values to "N/A"
	03		Gathering all other explainations Into one catagory "Yes"
+	04	Normalize values In "explaination"	More normalization for "explaination" values

UFO Reports Analysis (1950-2025)

Total Reports

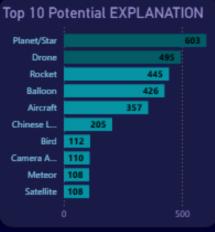
138K

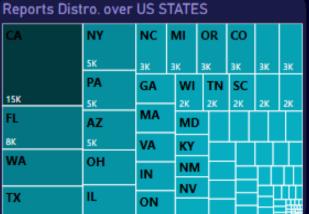






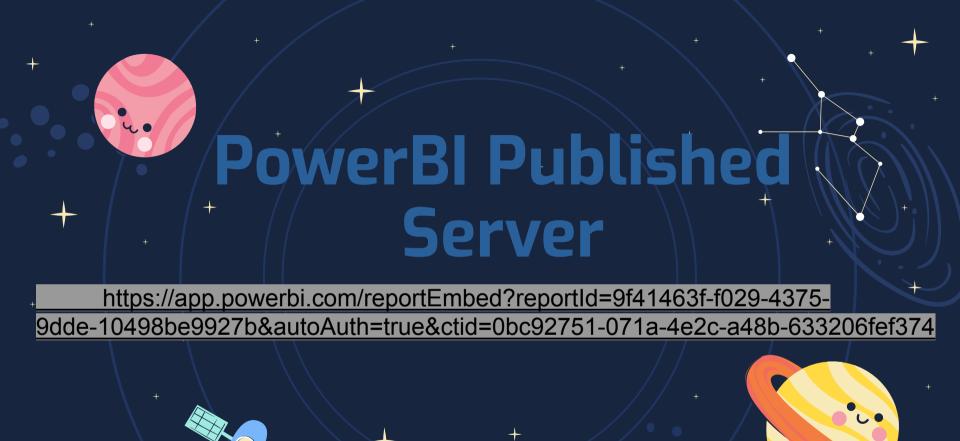
















A glowing white object passed silently across the sky, and suddenly a circular gate opened and it disappeared inside.

A star sped unbelievably fast, met another, vanished; then another came from a different angle, fast too-then disappeared.

2025-07-17 23:00:00

2025-07-08 21:16:00

Alexandria Governorate

Asyut



Some Insights

1. Peak reports Was between (2012:2014)

Insights

4. Most US states by reports num.

CA, FL and WL



94% of reports didn't got covered by media

3. Most common potential Explaination

Being a planet or a star

