

Open Brewery Data Lake Documentation

Project Overview

This project establishes a modern Data Lake using the Medallion Architecture (Bronze, Silver, Gold). The objective is to ingest, clean, and analyze global brewery data from the Open Brewery DB API to derive business insights.

Data Architecture

1. Bronze Layer (Raw)

- **Process**: Automated ingestion from Open Brewery DB via REST API.
- **Storage**: JSON format, preserving raw state.

2. Silver Layer (Cleaned)

- **Process**: Data cleaning, deduplication, and standardization.
- **Storage**: Optimized Parquet format partitioned by `brewery_type`.

3. Gold Layer (Analytical)

- **Process**: Complex aggregations and business metrics calculation.
- **Quality**: Automated checks verify data integrity.

Analytical Insights

1. Digital Maturity Score

- **Definition**: Percentage of breweries in a state that have both a registered website and phone number.
- **Insight**: High scores indicate a digitally advanced and reachable market.

2. Regional Diversity Index

- **Definition**: Count of unique brewery types present in the state.
- **Insight**: High variety indicates a heterogeneous and mature market.

3. Market Specialization (Micro Count)

- **Definition**: Total breweries vs. "micro" breweries per state.
- **Insight**: Highlights centers of craft beer production.

4. Data Trust Score

- **Definition**: Average completeness of Address, Phone, and Website.
- **Insight**: Indicator of business database reliability.

5. Geographic Hubs

- **Definition**: Cities with the highest concentration of breweries.

Open Brewery Data Lake - Relatório Estratégico

- ****Insight**:** Identifies competitive hotspots.

6. Breweries by Type and State

- ****Definition**:** Breakdown of brewery counts by category (micro, nano, brewpub, etc.) for each state.
- ****Insight**:** Useful for understanding the industrial profile of each region.

7. Global Distribution by Country and Type

- ****Definition**:** Comparison of brewery types across different countries.
- ****Insight**:** Reveals international trends in the brewing industry.

8. Geographic Coverage by State

- ****Definition**:** Deep dive into the distribution of breweries across cities within states.
- ****Insight**:** Essential for logistics and regional market penetration planning.

Open Brewery Data Lake - Relatório Estratégico

Detailed Data Tables (Top Results)

Breweries By Type And State

brewery_type	state_province	brewery_count
micro	ACT	7
micro	Alabama	26
closed	Alabama	15
brewpub	Alabama	5
planning	Alabama	3
taproom	Alabama	3
micro	Alaska	28
brewpub	Alaska	18
closed	Alaska	7
taproom	Alaska	4

Breweries By Country And Type

country	brewery_type	brewery_count
Australia	micro	491
Australia	large	22
Australia	regional	1
Austria	large	10
Austria	bar	2
Austria	nano	2
Austria	brewpub	1
Canada	micro	54
Canada	brewpub	44
Canada	closed	12

Top Cities By Brewery Count

city	state_province	country	brewery_count
Denver	Colorado	United States	92
San Diego	California	United States	91
Portland	Oregon	United States	85
Seattle	Washington	United States	85
Chicago	Illinois	United States	64
Austin	Texas	United States	49
Houston	Texas	United States	40
San Francisco	California	United States	39
Minneapolis	Minnesota	United States	39
Cincinnati	Ohio	United States	34

Geo Coverage By State

country	state_province	city	brewery_count
Australia	ACT	Fyshwick	4

Open Brewery Data Lake - Relatório Estratégico

country	state_province	city	brewery_count
Australia	ACT	Braddon	2
Australia	ACT	Gungahlin	1
Australia	NSW	Marrickville	12
Australia	NSW	Brookvale	6
Australia	NSW	Alexandria	4
Australia	NSW	Botany	4
Australia	NSW	South Murwillumbah	4
Australia	NSW	Armidale	3
Australia	NSW	Newtown	3

Digital Maturity

state_province	total_breweries	digitally_ready_count	maturity_score
ACT	7	7	100.0
Argyll	4	4	100.0
Chungcheongnamdo	3	3	100.0
Clare	1	1	100.0
Carlow	1	1	100.0
Busan	9	9	100.0
Bute	1	1	100.0
Bouche du Rhône	3	3	100.0
Bolzano	4	4	100.0
Kerry	3	3	100.0

Regional Diversity

state_province	unique_brewery_types
Washington	11
Wisconsin	9
Maryland	8
Texas	8
Oregon	8
Colorado	8
California	8
Missouri	8
Tennessee	8
Ohio	8

Market Specialization

state_province	total_brewery_count	micro_brewery_count
California	919	466
Washington	498	245
Colorado	449	228
New York	419	226
Michigan	375	143
Texas	352	196
Pennsylvania	345	129
Florida	312	190

Open Brewery Data Lake - Relatório Estratégico

state_province	total_brewery_count	micro_brewery_count
North Carolina	311	185
Ohio	303	163

Data Trust Score

state_province	trust_score
ACT	100.0
Argyll	100.0
Chungcheongnamdo	100.0
Clare	100.0
Carlow	100.0
Busan	100.0
Bute	100.0
Bouche du Rhône	100.0
Bolzano	100.0
Kerry	100.0