Data Dictionary

Index

1. [closed\_deals](#4vk0uou6dqkb)
2. [customers](#l3j5cphffmq2)
3. [geolocation](#l4vuhma176em)
4. [marketing\_qualified\_leads](#tupdggwiui1k)
5. [order\_items](#cwxdfsbxbpz)
6. [order\_payments](#d97x865br12q)
7. [order\_reviews](#ipeeumrrn3tc)
8. [orders](#7gduzg8utng8)
9. [product\_category\_name\_translation](#p8k34ornihqc)
10. [products](#5iyucymlrqxk)
11. [sellers](#3h722ipttgou)
12. [world\_bank\_latin\_america](#gxm56mvoacqf)
13. **close****d\_deals**

**General Data**

| **DataBase** | olist |
| --- | --- |
| **Classification** | Internal-customer-delivered |
| **Location** | [s3://olist-athena-query-result/tables/2e7c82b2-1a1c-4878-8d4c-aaa4e0a37484/](https://console.aws.amazon.com/s3/home?region=us-east-1&bucket=olist-athena-query-result&prefix=tables/2e7c82b2-1a1c-4878-8d4c-aaa4e0a37484/) |
| **Input Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat |
| **Output Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat |

**Table Detail**

| **Column Name** | **Description** | **Type** | **Indexes** | **Coments** |
| --- | --- | --- | --- | --- |
| mql\_id | Marketing Qualified Leads ID | string |  |  |
| seller\_id | Sellers ID | string |  |  |
| sdr\_id | No info | string |  |  |
| sr\_id | No info | string |  |  |
| won\_date | Close date | TimeStamp |  |  |
| business\_segment | Segment of Business | string |  |  |
| lead\_type | Type Lead | string |  |  |
| lead\_behaviour\_profile | Profile of behavior lead | string |  |  |
| business\_type | Type of Business | string |  |  |

1. **c****ustomers**

**General Data**

| **Database** | olist |
| --- | --- |
| **Classification** | Internal-customer-delivered |
| **Location** | [s3://olist-athena-query-result/tables/29155317-64c7-4803-9896-46ffbd185599/](https://console.aws.amazon.com/s3/home?region=us-east-1&bucket=olist-athena-query-result&prefix=tables/29155317-64c7-4803-9896-46ffbd185599/) |
| **Input Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat |
| **Output Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat |

**Table Detail**

| **Column Name** | **Description** | **Type** | **Indexes** | **Coments** |
| --- | --- | --- | --- | --- |
| customer\_id | Client ID | string |  |  |
| customer\_unique\_id | Corporation Client ID | string |  |  |
| customer\_zip\_code\_prefix | Zip Code | bigint |  |  |

1. **g****eolocation**

**General Data**

| **Database** | olist |
| --- | --- |
| **Classification** | Internal-customer-delivered |
| **Location** | [s3://olist-athena-query-result/tables/34123d52-6065-481d-9e6b-a6f969fad9d4/](https://console.aws.amazon.com/s3/home?region=us-east-1&bucket=olist-athena-query-result&prefix=tables/34123d52-6065-481d-9e6b-a6f969fad9d4/) |
| **Input Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat |
| **Output Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat |

**Table Detail**

| **Column Name** | **Description** | **Type** | **Indexes** | **Coments** |
| --- | --- | --- | --- | --- |
| zip\_code\_prefix | Zip Code | bigint | PK |  |
| latitude | latitude | double |  |  |
| longitude | longitude | double |  |  |
| city | city | string |  |  |
| state | state | string |  |  |

1. **marketing\_qualified\_leads**

**General Data**

| **Database** | olist |
| --- | --- |
| **Classification** | Internal-customer-delivered |
| **Location** | [s3://olist-athena-query-result/tables/e26bf960-e879-4c69-89ee-fe4f17c63733/](https://console.aws.amazon.com/s3/home?region=us-east-1&bucket=olist-athena-query-result&prefix=tables/e26bf960-e879-4c69-89ee-fe4f17c63733/) |
| **Input Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat |
| **Output Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat |

**Table Detail**

| **Column Name** | **Description** | **Type** | **Indexes** | **Coments** |
| --- | --- | --- | --- | --- |
| mql\_id | Marketing Qualified leads | string | PK |  |
| first\_contact\_date | First contact date | Timestamp |  |  |
| landing\_page\_id | Landing page id | string |  |  |
| origin | origin | string |  |  |

1. **order\_items**

**General Data**

| **Database** | [olist](https://us-east-1.console.aws.amazon.com/glue/home?region=us-east-1#/v2/data-catalog/databases/view/olist?catalogId=781520571669) |
| --- | --- |
| **Classification** | Internal-customer-delivered |
| **Location** | [s3://olist-athena-query-result/tables/001bc5f9-6522-4664-a6d8-27859ec7140f/](https://console.aws.amazon.com/s3/buckets/olist-athena-query-result/tables/001bc5f9-6522-4664-a6d8-27859ec7140f/?region=us-east-1) |
| **Input Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat |
| **Output Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat |

**Table Detail**

| **Column Name** | **Description** | **Type** | **Indexes** | **Coments** |
| --- | --- | --- | --- | --- |
| order\_id | Order id | string | PK |  |
| order\_item\_id | Order ítem id | bigint | PK | - |
| product\_id | Product id | string | - | - |
| seller\_id | Seller id | string | - | - |
| shipping\_limit\_date | Shipping limit date | Timestamp | - | - |
| price | Price | double | - | - |
| freight\_value | freight value | double |  |  |

1. **order\_payments**

**General Data**

| **Database** | olist |
| --- | --- |
| **Classification** | Internal-customer-delivered |
| **Location** | [s3://olist-athena-query-result/tables/bc9742a1-b729-410a-8b23-fa5b336e03c9/](https://console.aws.amazon.com/s3/home?region=us-east-1&bucket=olist-athena-query-result&prefix=tables/bc9742a1-b729-410a-8b23-fa5b336e03c9/) |
| **Input Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat |
| **Output Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat |

**Table Detail**

| **Column Name** | **Description** | **Type** | **Indexes** | **Coments** |
| --- | --- | --- | --- | --- |
| order\_id | Order ID | string | PK |  |
| payment\_sequential | payment\_sequential | bigint |  |  |
| payment\_type | payment\_type | string |  |  |
| payment\_installments | payment\_installments | bigint |  |  |
| payment\_value | payment\_value | double |  |  |

1. **order\_reviews**

**General Data**

| **Database** | olist |
| --- | --- |
| **Classification** | Internal-customer-delivered |
| **Location** | [s3://olist-athena-query-result/tables/23f526d2-f831-4389-a6fe-de4204116ef8/](https://console.aws.amazon.com/s3/home?region=us-east-1&bucket=olist-athena-query-result&prefix=tables/23f526d2-f831-4389-a6fe-de4204116ef8/) |
| **Input Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat |
| **Output Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat |

**Table Detail**

| **Column Name** | **Description** | **Type** | **Indexes** | **Coments** |
| --- | --- | --- | --- | --- |
| review\_id | Reviews or comments ID | string | PK |  |
| order\_id | Order ID | string |  |  |
| review\_score | Comments score | bigint |  |  |
| review\_comment\_message | Review comment message | string |  |  |
| review\_creation\_date | Review creation date | Timestamp |  |  |
| review\_answer\_timestamp | review\_answer\_timestamp | Timestamp |  |  |

1. **orders**

**General Data**

| **Database** | olist |
| --- | --- |
| **Classification** | Internal-customer-delivered |
| **Location** | [s3://olist-athena-query-result/tables/3bf08c73-9299-48da-8bcb-521a0cc46e06/](https://console.aws.amazon.com/s3/home?region=us-east-1&bucket=olist-athena-query-result&prefix=tables/3bf08c73-9299-48da-8bcb-521a0cc46e06/) |
| **Input Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat |
| **Output Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat |

**Table Detail**

| **Column Name** | **Description** | **Type** | **Indexes** | **Coments** |
| --- | --- | --- | --- | --- |
| order\_id | Order ID | string | PK |  |
| customer\_id | Customer ID | string |  |  |
| order\_status | Status order | string |  |  |
| order\_purchase\_timestamp | Purchase order date | Timestamp |  |  |
| order\_approved\_at | Approved order date | Timestamp |  |  |
| order\_delivered\_carrier\_date | Order delivered carrier date | Timestamp |  |  |
| order\_delivered\_customer\_date | Order delivered Customer date | Timestamp |  |  |
| order\_estimated\_delivery\_date | review\_answer\_timestamp | Timestamp |  |  |

1. **product\_category\_name\_translation**

**General Data**

| **Database** | olist |
| --- | --- |
| **Classification** | Internal-customer-delivered |
| **Location** | [s3://olist-athena-query-result/tables/88731f82-761e-45b7-97ca-013f377f2b3f/](https://console.aws.amazon.com/s3/home?region=us-east-1&bucket=olist-athena-query-result&prefix=tables/88731f82-761e-45b7-97ca-013f377f2b3f/) |
| **Input Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat |
| **Output Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat |

**Table Detail**

| **Column Name** | **Description** | **Type** | **Indexes** | **Coments** |
| --- | --- | --- | --- | --- |
| product\_category\_id | Category Product ID | bigint | PK |  |
| product\_category\_name | Name of category product in portuguese | string |  |  |
| product\_category\_name\_english | Name of category product in english | string |  |  |

1. **p****roducts**

**General Data**

| **Database** | olist |
| --- | --- |
| **Classification** | Internal-customer-delivered |
| **Location** | [s3://olist-athena-query-result/tables/ed06dba7-5191-490a-a9f3-54f627d9e041/](https://console.aws.amazon.com/s3/home?region=us-east-1&bucket=olist-athena-query-result&prefix=tables/ed06dba7-5191-490a-a9f3-54f627d9e041/) |
| **Input Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat |
| **Output Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat |

**Table Detail**

| **Column Name** | **Description** | **Type** | **Indexes** | **Coments** |
| --- | --- | --- | --- | --- |
| product\_id | Product id | string | PK |  |
| product\_category\_ID | Product category ID | string |  |  |
| product\_name\_length | Product name length | bigint |  |  |
| product\_description\_length | Product Description length | bigint |  |  |
| product\_photos\_qty | Product photos qty | bigint |  |  |
| product\_weight\_g | Product weight g | bigint |  |  |
| product\_length\_cm | Product length cm | bigint |  |  |
| product\_height\_cm | Product height cm | bigint |  |  |
| product\_width\_cm | Product width cm | bigint |  |  |

1. **s****ellers**

**General Data**

| **Database** | olist |
| --- | --- |
| **Classification** | Internal-customer-delivered |
| **Location** | [s3://olist-athena-query-result/tables/a2b32a33-ba82-4a2e-82a6-6df44b6552fb/](https://console.aws.amazon.com/s3/home?region=us-east-1&bucket=olist-athena-query-result&prefix=tables/a2b32a33-ba82-4a2e-82a6-6df44b6552fb/) |
| **Input Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat |
| **Output Format** | org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat |

**Table Detail**

| **Column Name** | **Description** | **Type** | **Indexes** | **Coments** |
| --- | --- | --- | --- | --- |
| seller\_id | Seller ID | string | PK |  |
| seller\_zip\_code\_prefix | Zip Code del Seller | bigint |  |  |

1. **w****orld\_bank\_latin\_america**

**General Data**

| **Database** | olist |
| --- | --- |
| **Classification** | Obtained from the World Bank source |
| **Location** | [s3://olist-processed-useast1-dev/worldbankLatinAmerica-job\_17Jan2023\_1673981810939/](https://console.aws.amazon.com/s3/home?region=us-east-1&bucket=olist-processed-useast1-dev&prefix=worldbankLatinAmerica-job_17Jan2023_1673981810939/) |
| **Input Format** | org.apache.hadoop.mapred.TextInputFormat |
| **Output Format** | org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat |

**Table Detail**

| **Column Name** | **Description** | **Type** | **Indexes** | **Coments** |
| --- | --- | --- | --- | --- |
| país name | Country | string |  |  |
| país code | Country code | string | PK |  |
| serie name | serie name | string |  |  |
| serie code | serie code | string | PK |  |
| 2017 [yr2017] | 2017 values | double |  |  |
| 2018 [yr2018] | 2018 values | double |  |  |
| 2019 [yr2019] | 2019 values | double |  |  |
| 2020 [yr2020] | 2020 values | double |  |  |
| 2021 [yr2021] | 2021 values | double |  |  |