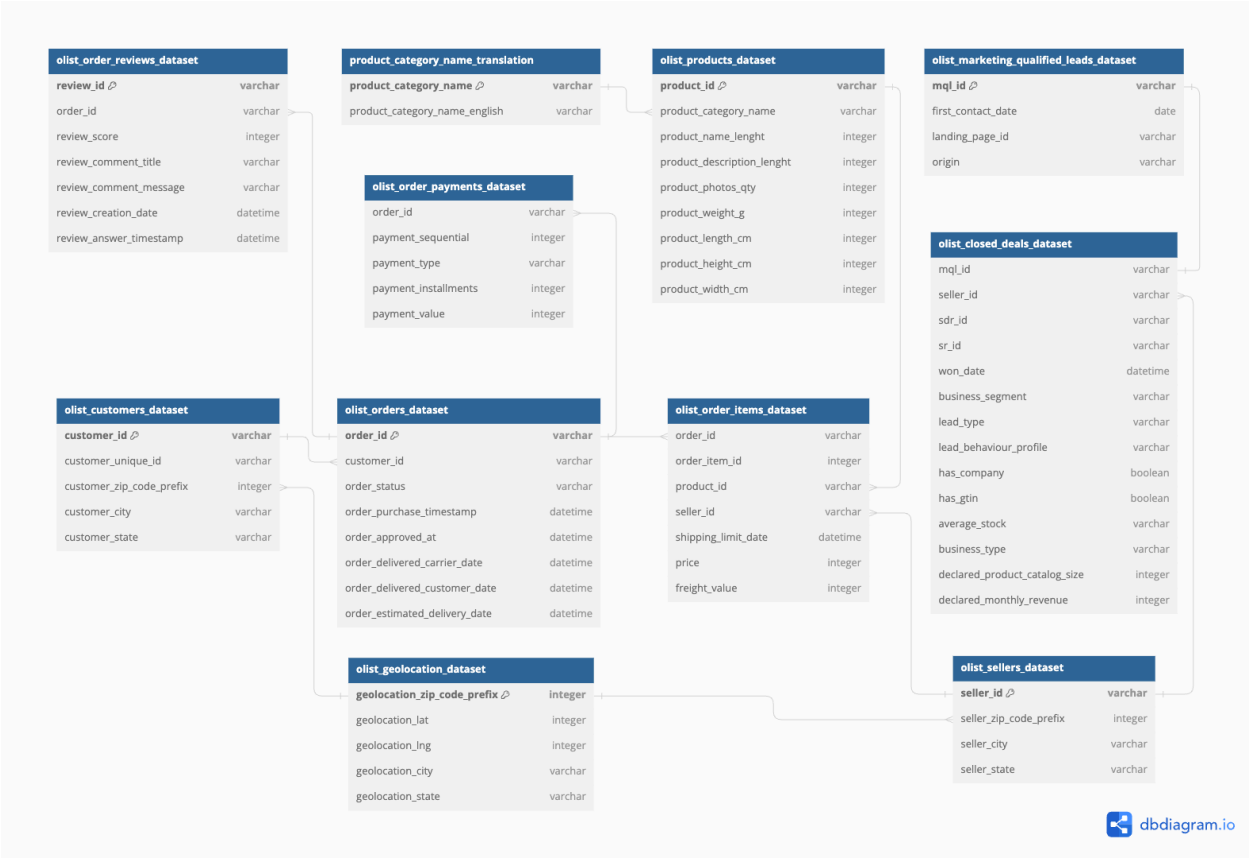
Olist Sales Analysis Technical Documentation

Tools & Environment

- For the Olist Sales Analysis Project, I used **BigQuery** as my database and utilized **SQL** for data cleaning, exploration, and manipulation.
- For data visualization, I used **Tableau Public**.
- The Olist dataset contains 11 tables. For the purposes of the project, I only use four tables namely: **orders, order_items, products,** and **product_category name_translation**. The full ERD of the dataset is found below:



SQL Queries

Query #1: JOIN relevant tables

- Create a raw table joining all relevant tables from the dataset. The resulting table serves as the order_sales_data table that we use all throughout the project.
- This raw table serves as a backup table as well.

```
CREATE OR REPLACE TABLE `olist_dataset.order_sales_data_raw` AS
SELECT a.order_id, a.order_item_id, a.product_id, b.category, a.order_purchase_timestamp,
a.price
FROM (
  SELECT a.order_id, a.order_item_id, a.product_id, b.order_purchase_timestamp, a.price
  FROM `olist_dataset.order_items` a
  LEFT JOIN `olist_dataset.orders` b
  ON a.order_id = b.order_id
  ) a
LEFT JOIN (
  SELECT a.product_id, b.string_field_1 AS category
  FROM `olist_dataset.products` a
  LEFT JOIN `olist_dataset.product_category_name_translation` b
  ON a.product_category_name = b.string_field_0
  ) b
ON a.product_id = b.product_id
GROUP BY a.order_id, a.order_item_id, a.product_id, b.category, a.order_purchase_timestamp,
a.price
ORDER BY a.order_purchase_timestamp ASC, a.order_item_id ASC;
```

Query #2: Data cleaning

- From the raw table created, create a new clean table. In the query below, we deal with missing values in the ‘category’ column and then create a new clean table.

```
CREATE OR REPLACE TABLE `olist_dataset.order_sales_data` AS
SELECT DATE(order_purchase_timestamp) AS order_date, COALESCE(category, 'Unknown') AS
product_category,
FROM `olist_dataset.order_sales_data_raw`;
```

Query #3: Manipulate, Aggregate, Perform Operations

- We then create an order_ total_sales table where we manipulate specific columns such as the order_purchase_timestamp and the product_category. We then aggregate the product_category to create the fields: quantity_sold and total_sales.

```sql
-- SQL query that creates the total_sales table
CREATE OR REPLACE TABLE `olist_dataset.order_total_sales` AS
SELECT
  order_id,
  product_id,
  DATE(order_purchase_timestamp) AS order_date,
  LOWER(product_category) AS product_category,
  COUNT(product_category) AS quantity_sold,
  price,
  COUNT(product_category) * price AS total_sales
FROM
  `olist_dataset.order_sales_data`
GROUP BY
  order_id,
  product_id,
  order_date,
  product_category,
  price
ORDER BY
  order_date, product_category;
```

Query #4: Add fields to the table

- Now that we have information on quantity_sold and total_sales, we can now compute the Average Order Value (AOV) which is a key metric for our analysis.

```sql
-- SQL query that creates avg_order_value table
CREATE OR REPLACE TABLE `olist_dataset.order_avg_order_value` AS
SELECT
    order_date,
    SUM(total_sales) / COUNT(DISTINCT order_id) AS avg_order_value
FROM
    `olist_dataset.order_total_sales`
GROUP BY
    order_date
ORDER BY
    order_date;
```

Query #5: Finalize the table

- Finalize the table that will be exported from BigQuery and then imported to Tableau Public for data visualizations and dashboarding.

```sql
CREATE OR REPLACE TABLE `olist_dataset.order_sales_avg_value` AS
SELECT a.order_id, a.product_id, a.order_date, a.product_category, a.quantity_sold, a.price,
a.total_sales, b.avg_order_value
FROM `olist_dataset.order_total_sales` a
JOIN `olist_dataset.order_avg_order_value` b
ON a.order_date = b.order_date
ORDER BY a.order_date;
```

Query #6: Pull the final table

- Pull the final table, export as a csv file, and import to Tableau for visualization.

```
-- Pulls the order_sales_avg_value table to be exported
SELECT *
FROM `olist_dataset.order_sales_avg_value`;
```

Conclusion

- The final resulting table contains all the relevant information for our analysis. A glimpse of the final dataset is found below:

olist_orders_sales_data

File Edit View Insert Format Data Tools Extensions Help

J1

	A	B	C	D	E	F	G	H
1	order_id	product_id	order_date	product_categor	quantity_sold	price	total_sales	avg_order_value
2	2e7a8482f6fb09	c1488892604e4	2016-09-04	furniture_decor	1	39.99	39.99	72.89
3	2e7a8482f6fb09	f293394c72c9b5	2016-09-04	furniture_decor	1	32.9	32.9	72.89
4	e5fa5a7210941f	f3c2d01a84c947	2016-09-05	telephony	1	59.5	59.5	59.5
5	bfbd0f9bdef843	5a6b04657a4c5	2016-09-15	health_beauty	3	44.99	134.97	134.97
6	71303d7e93b39	d2998d7ced12f8	2016-10-02	baby	1	100	100	100
7	cd3b8574c82b4	e2a1d45a73dc7	2016-10-03	fashion_shoes	1	29.99	29.99	57.935
8	d207cc2726756	107177bf61755f	2016-10-03	furniture_decor	1	119.9	119.9	57.935
9	ef1b29b591d31	bfce5e847034e1	2016-10-03	furniture_decor	1	74.9	74.9	57.935
10	be5bc2f0da14d8	fd7fd78fd3cbc1b	2016-10-03	sports_leisure	1	21.9	21.9	57.935
11	65d1e226dfaeb8	9c7bdf67b06b41	2016-10-03	sports_leisure	1	21.5	21.5	57.935
12	a41c8759fbe7a	b72b39418216e	2016-10-03	sports_leisure	1	36.49	36.49	57.935
13	ae8a60e4b03c5	a5c3ddb1a400f5	2016-10-03	toys	1	128.9	128.9	57.935
14	3b697a20d9e42	3ae08df6bcbfe2	2016-10-03	watches_gifts	1	29.9	29.9	57.935
15	79ffdd52a918bb	5691a536d459e	2016-10-04	air_conditioning	1	299.99	299.99	165.6826667
16	893dc49276696	27051f70cfcbd6	2016-10-04	air_conditioning	1	149.9	149.9	165.6826667
17	dd11631d8b027	27051f70cfcbd6	2016-10-04	air_conditioning	2	149.9	299.8	165.6826667
18	c4b41c36dd589	9e93b2c4cb5ee	2016-10-04	auto	1	9.9	9.9	165.6826667
19	bb9fd560a08ca7	3c37a8a5d5ef2e	2016-10-04	baby	1	199.99	199.99	165.6826667
20	8b388d845a3e8	35bd74b6319ba	2016-10-04	baby	1	39.9	39.9	165.6826667
21	8b388d845a3e8	dc82d7e8f9996	2016-10-04	baby	1	39.9	39.9	165.6826667
22	bcbc0973414b1	48a12f318432c4	2016-10-04	bed_bath_table	1	79.9	79.9	165.6826667
23	bcbc0973414b1	66e08854f9319	2016-10-04	bed_bath_table	1	79.9	79.9	165.6826667
24	dd359d3c29445	f9bcf9d8e89d01	2016-10-04	computers_acce	1	164	164	165.6826667
25	5cc475c7c0329	3410cbd7df213	2016-10-04	computers_acce	1	71	71	165.6826667

- This dataset can be accessed here.
- The Tableau dashboard created for this project can be found here.