

Performance & Analytics Dashboard

BUSINESS INTELLIGENCE REPORT

MEYER-PETGRAVE, FITZROY – A0384858
YERIDE UNIVERSITY | ENGLAND

Table of Contents

| | |
|--|------------------------------|
| LIST OF FIGURES | ERROR! BOOKMARK NOT DEFINED. |
| A. SECTION 1: BUSINESS INTELLIGENCE DESIGN | 3 |
| 1 DATA SOURCE DESCRIPTION AND BUSINESS QUESTIONS | 3 |
| 1.1 OLIST MARKETPLACE DATASET..... | 3 |
| 1.1.1 <i>Customer dataset</i> | 4 |
| 1.1.2 <i>Geolocation dataset</i> | 4 |
| 1.1.3 <i>Order Items Dataset</i> | 5 |
| 1.1.4 <i>Order Payment Dataset</i> | 5 |
| 1.1.5 <i>Order reviews dataset</i> | 6 |
| 1.1.6 <i>Order Dataset</i> | 6 |
| 1.1.7 <i>Products Dataset</i> | 7 |
| 1.1.8 <i>Sellers Dataset</i> | 7 |
| 1.1.9 <i>Category Name Translation</i> | 8 |
| 1.2 BRAZILIAN STATES INFORMATION DATASET | 8 |
| 1.3 SUMMARY OF DATASET | 9 |
| 1.4 KEY PERFORMANCE INDICATORS..... | 12 |
| 1.5 BUSINESS QUESTIONS | 12 |
| 1.6 USER GROUP FOR THIS REPORT | 12 |
| 1.7 WHY IS THIS INFORMATION NEEDED? | 13 |
| 2 DATA PRE-PROCESSING AND DATA CLEANSING..... | 13 |
| 2.1 IMPORTING DATASETS..... | 13 |
| 2.1.1 <i>Import Olist Datasets from Excel into Power Query</i> | 13 |
| 2.1.2 <i>Use First Rows as Headers</i> | 15 |
| 2.1.3 <i>Import Brazil Dataset into Power Query from the Web</i> | 16 |
| 2.2 DATA CLEANING AND PREPARATION..... | 17 |
| 2.2.1 <i>Delete Existing Relationships</i> | 18 |
| 2.2.2 <i>Remove Null Values</i> | 18 |
| 2.2.3 <i>Remove Duplicates</i> | 20 |
| 2.2.4 <i>Join tables using the 'Merge Queries' function</i> | 20 |
| 2.2.5 <i>Delete Columns</i> | 21 |
| 2.2.6 <i>Replace Values</i> | 22 |
| 2.2.7 <i>Change Data Type</i> | 23 |
| 2.2.8 <i>Split columns by Delimiter</i> | 24 |
| 2.2.9 <i>Remove Column</i> | 25 |
| 2.2.10 <i>Rename Column</i> | 26 |
| 2.2.11 <i>Rename Tables</i> | 27 |
| 3 DATA MODELLING – STAR SCHEMA..... | 29 |

| | | |
|-----------|--|------------------------------|
| 3.1 | FACT TABLES | ERROR! BOOKMARK NOT DEFINED. |
| 3.2 | DIMENSION TABLES..... | ERROR! BOOKMARK NOT DEFINED. |
| 3.3 | DELETE RELATIONSHIP | 29 |
| 3.4 | CREATE NEW RELATIONSHIPS | 31 |
| 3.5 | CREATE DATE TABLE USING DAX..... | 34 |
| B. | SECTION 2: BUSINESS INTELLIGENCE REPORT | 36 |
| 3.6 | EXECUTIVE SUMMARY..... | 37 |
| 3.7 | INTRODUCTION | 38 |
| 3.7.1 | <i>Key Performance Indicators</i> | 38 |
| 3.7.2 | <i>Business Questions Addressed</i> | 38 |
| 3.7.3 | <i>Description of The Data Used in The Model</i> | 38 |
| 3.8 | FINDING BASED ON ANALYSIS AND EVALUATION (MEASURES/TABLES/COLUMNS) | 39 |
| 3.8.1 | <i>Performance Analytics</i> | 39 |
| 3.8.1.1 | Variance Measure | 39 |
| 3.8.1.2 | Year-Over-Year Growth Rate Measure (Sales and Orders)..... | 40 |
| 3.8.1.3 | Month over month percentage growth(measure) – sales and orders | 41 |
| 3.8.2 | <i>Current Insights (Analysing Customer Satisfaction)</i> | 42 |
| 3.8.2.1 | Satisfaction Mapping Based on Ratings (M Language in Power Query)..... | 42 |
| 3.8.2.2 | Ranking Category Table (Using M Language in Power Query) | 42 |
| 3.8.2.3 | Percentage Satisfaction Measure (Using DAX) | 43 |
| 3.8.2.4 | Category Mapping – new column (DAX) | 44 |
| 3.8.3 | <i>New Business Drivers – Discovering KPI Influencers</i> | 45 |
| 3.8.3.1 | Product Volume (Custom column)..... | 45 |
| 3.8.3.2 | Population density (Custom Column) | 45 |
| 3.9 | FINDING BASED ON ANALYSIS AND EVALUATION (CHARTS/GRAPHS)..... | 46 |
| 3.9.1 | <i>Performance Overview</i> | 46 |
| 3.9.2 | <i>Sales Analytics</i> | 47 |
| 3.9.2.1 | What are the trends in sales? | 47 |
| 3.9.2.2 | How have the Sales changed over time? | 48 |
| 3.9.3 | <i>Orders Analytics</i> | 48 |
| 3.9.3.1 | What are the trends in number of orders? | 49 |
| 3.9.3.2 | How have the Orders changed over time? | 49 |
| 3.9.4 | <i>Dynamic Drill-through</i> | 50 |
| 3.9.4.1 | Which locations and products generated the most revenue and orders? | 50 |
| 3.9.5 | <i>Key Influencers</i> | 51 |
| 3.9.5.1 | How are the KPSs affected by Product and Location Parameters? | 51 |
| 3.10 | CONCLUSIONS AND RECOMMENDATIONS | 52 |
| 3.10.1 | <i>Summary of Report</i> | 52 |
| 3.10.2 | <i>Recommendations based on findings</i> | 52 |
| 3.10.3 | <i>Personal Conclusions</i> | 53 |

A. SECTION 1: BUSINESS INTELLIGENCE DESIGN

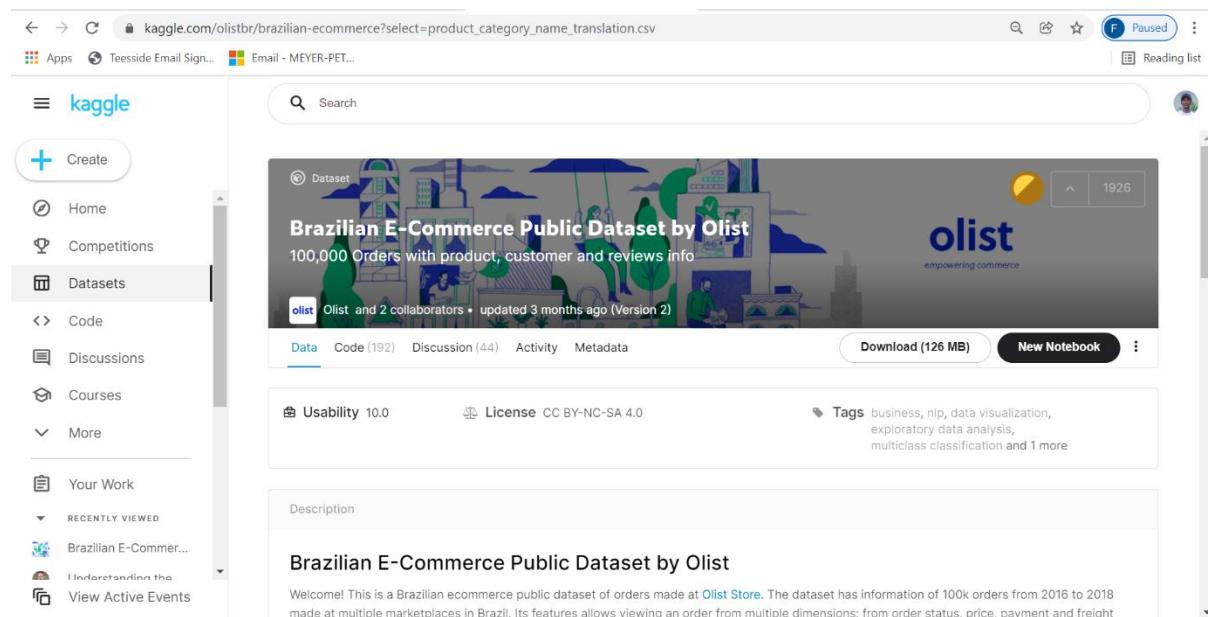
1 DATA SOURCE DESCRIPTION AND BUSINESS QUESTIONS

1.1 Olist Marketplace Dataset

The primary dataset used for this business intelligence design is sales data internally generated from Olist E-commerce platform from 2016 to 2018. This dataset was released by Olist on the Kaggle platform for the purpose of creating solutions to business problems. The dataset - containing over 100,000 rows, 9 tables and 52 columns - was sourced from the link below as shown in figure 1.

<https://www.kaggle.com/olistbr/brazilian-e-commerce>

Figure 1: Olist Marketplace Dataset on Kaggle



The Olist platform functions by connecting small businesses to larger product marketplaces to help entrepreneurs sell their products to a larger customer base. The sellers list products on the platform and the buyers place orders, Olist delivery partner fulfills the order and Olist makes a percentage of the revenue.

After a customer purchases the product from Olist Store a seller gets notified to fulfil that order. Once the customer receives the product, or the estimated delivery date is due, the customer gets a satisfaction survey by email where he can give a note for the purchase experience and write down some comments.

After an order is completed, the platform retains details about the orders, customers, sellers, location, delivery, and reviews. Below is a snapshot of the order_items table.

1.1.1 Customer dataset

This dataset has information about the customer and its location. Use it to identify unique customers in the orders dataset and to find the orders delivery location.

Figure 2: Customer dataset viewed in excel

| customer_id | customer_unique_id | customer_zip_code_prefix | customer_city | customer_state |
|----------------------------------|-----------------------------------|--------------------------|-----------------------|----------------|
| 06b8999e2fba1a1fb8172c00ba8c7 | 861ef4711a0842e4b93843c6dd7febb0 | 14409 | franca | SP |
| 18955e83d337fd62de6f018a428ac77 | 290c77bc52967a935b93aa66c33dc3 | 9790 | sao bernardo do campo | SP |
| 4e7b3a00288586ebd08712fd0374a03 | 060e732b5b29e8181a18229c7b0b2b5e | 1151 | sao paulo | SP |
| b2b6027bc5109e5294ddc6358b12c3 | 259dac757896d24d77029acbbff3fc | 8775 | mogi das cruzes | SP |
| 4f2d8ab171c80e83647c12e35b23ad | 345ecd01c38a18a9036ed96c73b8d066 | 13056 | campinas | SP |
| 7879864db9bc304752a9c2b2e1212b8 | 4c93744516667d3d8b1f645a3116a4 | 89254 | jaragua do sul | SC |
| fd826e7f63160e536e9908c76c3f441 | addec96d2e059c8030f6871d30d177 | 4534 | sao paulo | SP |
| 5e274a7a0c3809e14aba7ad5aae0d407 | 5752a98a409812fe9618067b6b8bebe4f | 35182 | timoteo | MG |
| 5ad0f8e34b2e993982a47070956c565 | 1175e95fb47ddff9de6b2b061887e0d | 81560 | curitiba | PR |
| 4b7139f34592b3a31687243a302fa75b | 9afe194fb833f79e300e37e580171f22 | 30575 | belo horizonte | MG |
| 9fb35e4ed6f0a14a977cd94ea4042bb | 2a7745e1ed5160289ed9b29c7d0539a5 | 39400 | montes claros | MG |
| 5aa9e4fd4df020959cad2d772509598 | 2a46fb94ae5fcbbe850418118ce0e090 | 20231 | rio de janeiro | RJ |
| b2d1536598b73a9abd18e0d75d92f0a3 | 918d8c7cd72cd9f6ed4b4d42ed785235 | 18682 | lencois paulista | SP |
| eabebad39a88bb6f5b52376faec28612 | 295c05e81917928d76245e842748184d | 5704 | sao paulo | SP |
| 11fc7bf19b0401b292af6c1d4470b753 | 3151a81801c83863616b227747f5ecf | 95110 | caxias do sul | RS |
| 206f3129c04d7d0b9550426023f0a08 | 21f748a1614e1688a9014eb3ee6fa325 | 13412 | piracicaba | SP |
| a7c125a0a7b75146167bf704a7f8e98 | 5c29911bd08bbff3c410713e4de5a0b5 | 22750 | rio de janeiro | RJ |
| c56c1596a3bb6b0d0ce5766992c48a91 | b6e995616ef34a5b6d7ad92f8ed775 | 7124 | guarulhos | SP |
| 98b8ce803689b3562defaa4613ef426f | 7f1a72e8988ce735ba118d5f4147458 | 5416 | sao paulo | SP |
| 21490ea0986e0de72da77f15456a0e0 | 3ef6fb62fd4d99456a6a6820a40f2d79 | 68485 | pacaja | PA |
| 22154c4ded6991ba3cd249d11abf4130 | e607ede0e63436308660236f5a52d5e | 88034 | florianopolis | SC |
| 690172ab319622688d3b4df2f676898 | 2965dcfa0d3181817e2b946f921ea021 | 74914 | aparecida de goiania | GO |
| 293812140a20953c43ca8c98787fc | 482441ea6a06b1f7fe97847560ea75 | 5713 | sao paulo | SP |

1.1.2 Geolocation dataset

This dataset has information about Brazilian zip codes and its latitude and longitude coordinates.

Figure 3: Geolocation dataset viewed in excel

| A1 | B | C | D | E |
|-----------------------------|-----------------|-----------------|------------------|-------------------|
| geolocation_zip_code_prefix | geolocation_lat | geolocation_lng | geolocation_city | geolocation_state |
| 1037 | -23.54562128 | -46.63929205 | sao paulo | SP |
| 1046 | -23.54608113 | -46.6448203 | sao paulo | SP |
| 1041 | -23.54612897 | -46.64295148 | sao paulo | SP |
| 1041 | -23.54439216 | -46.63949931 | sao paulo | SP |
| 1031 | -23.54157796 | -46.64160722 | sao paulo | SP |
| 1012 | -23.5477623 | -46.63536054 | s o paulo | SP |
| 1047 | -23.54627311 | -46.6412517 | sao paulo | SP |
| 1013 | -23.54692321 | -46.6342637 | sao paulo | SP |
| 1029 | -23.54376906 | -46.63427784 | sao paulo | SP |
| 1011 | -23.54736955 | -46.63603162 | sao paulo | SP |
| 1013 | -23.54732513 | -46.63418379 | sao paulo | SP |
| 1032 | -23.5384181 | -46.6337838 | sao paulo | SP |
| 1014 | -23.54643534 | -46.63383023 | sao paulo | SP |
| 1012 | -23.54894599 | -46.63467113 | sao paulo | SP |
| 1037 | -23.54518734 | -46.63785524 | s o paulo | SP |
| 1041 | -23.54608113 | -46.6448203 | sao paulo | SP |
| 1039 | -23.54188301 | -46.63991947 | sao paulo | SP |
| 1024 | -23.54138952 | -46.62989909 | s o paulo | SP |
| 1009 | -23.5469354 | -46.63658793 | sao paulo | SP |
| 1046 | -23.54588428 | -46.64316319 | sao paulo | SP |
| 1046 | -23.54516589 | -46.6437858 | sao paulo | SP |
| 1020 | -23.55223547 | -46.62844066 | s o paulo | SP |
| 1011 | -23.54669049 | -46.63544696 | s o paulo | SP |

1.1.3 Order Items Dataset

This dataset includes data about the items purchased within each order

Figure 4: Order items dataset viewed in excel

| A1 | B | C | D | E | F | G | H |
|----------|------------------------------------|--------------|--------------------------|-------------------|------------------------------|-------------------------------|-------------------------------|
| order_id | customer_id | order_status | order_purchase_timestamp | order_approved_at | order_delivered_carrier_date | order_delivered_customer_date | order_estimated_delivery_date |
| 2 | e481f51cbcd54678b7cc49136fd2d6af7 | delivered | 02/10/2017 10:56 | 02/10/2017 11:07 | 10/10/2017 21:25 | 18/10/2017 00:00 | |
| 3 | 53cd2b1e8bc7dcde0b56741e2150273451 | delivered | 24/07/2017 20:41 | 26/07/2018 03:24 | 07/08/2018 15:27 | 13/08/2018 00:00 | |
| 4 | 47770e9100c2d0c449469cf07ec5577 | delivered | 08/08/2018 08:38 | 08/08/2018 08:55 | 08/08/2018 13:50 | 17/08/2018 00:00 | |
| 5 | 949d5b44dbfb5de918fe9c16f97b458a | delivered | 18/11/2017 19:28 | 18/11/2017 19:45 | 22/11/2017 13:39 | 02/12/2017 00:28 | 15/12/2017 00:00 |
| 6 | a621c59c840e6e0cb839c9eb557f8159 | delivered | 13/02/2018 21:18 | 13/02/2018 22:20 | 14/02/2018 19:46 | 16/02/2018 18:17 | 26/02/2018 00:00 |
| 7 | a6591c26518c18c1dec52889c2d48acc3 | delivered | 09/07/2017 21:57 | 09/07/2017 22:10 | 11/07/2017 14:58 | 26/07/2017 10:57 | 01/08/2017 00:00 |
| 8 | 126c7cefa42fd92c2ed53f7a9d7737a | invoiced | 11/04/2017 12:22 | 13/04/2017 13:23 | 16/05/2017 10:07 | 26/05/2017 12:55 | 09/05/2017 00:00 |
| 9 | 6514bbad0028c912c2374de245783f | delivered | 16/05/2017 13:10 | 16/05/2017 13:22 | 22/05/2017 10:07 | 26/05/2017 12:55 | 07/06/2017 00:00 |
| 10 | 76c6866289121a93b815b4852a3-3 | delivered | 23/01/2017 19:29 | 25/01/2017 02:50 | 26/01/2017 14:16 | 02/02/2017 14:08 | 06/03/2017 00:00 |
| 11 | e69fb5b88e0ed9a785585b27e16d8 | delivered | 29/07/2017 11:55 | 29/07/2017 12:05 | 10/08/2017 19:45 | 16/08/2017 17:14 | 23/08/2017 00:00 |
| 12 | e6c16c5b9ec1d901a0985e6118ae8 | delivered | 16/05/2017 19:41 | 16/05/2017 19:50 | 18/05/2017 11:40 | 29/05/2017 11:18 | 07/06/2017 00:00 |
| 13 | 34513c4f4f4ca62a5569c09ec7c7ed9 | delivered | 13/07/2017 19:58 | 13/07/2017 20:10 | 14/07/2017 18:43 | 19/07/2017 14:04 | 08/08/2017 00:00 |
| 14 | 83351660a693215100d6e95a6d32918 | delivered | 07/06/2018 10:06 | 08/06/2018 10:13 | 11/06/2018 13:29 | 19/06/2018 12:05 | 18/07/2018 00:00 |
| 15 | 5f99c150b717acbad13a77225a50 | delivered | 25/07/2018 17:44 | 25/07/2018 17:55 | 26/07/2018 13:16 | 30/07/2018 15:52 | 08/08/2018 00:00 |
| 16 | 432aa1f21d85167c2c86cc448442cc | delivered | 01/03/2018 14:14 | 01/03/2018 15:10 | 02/03/2018 21:09 | 12/03/2018 23:36 | 21/03/2018 00:00 |
| 17 | dc63651111c05097dc505de843-3 | delivered | 07/06/2018 19:03 | 12/06/2018 23:31 | 11/06/2018 14:54 | 21/06/2018 15:34 | 04/07/2018 00:00 |
| 18 | 40397836b0c04a22354-f531026e5f | delivered | 02/01/2018 19:00 | 02/01/2018 19:09 | 03/01/2018 18:19 | 20/01/2018 01:38 | 06/02/2018 00:00 |
| 19 | 1161009343049556ba9d35be0cd | delivered | 26/12/2017 23:41 | 26/12/2017 23:50 | 28/12/2017 18:33 | 08/01/2018 22:36 | 29/01/2018 00:00 |
| 20 | 85ce859f6dc634de62f2129044043 | delivered | 21/11/2017 00:03 | 21/11/2017 00:14 | 23/11/2017 21:32 | 27/11/2017 18:28 | 11/12/2017 00:00 |
| 21 | 831098c1e708f7b4e926f72fb889c2d7f | delivered | 26/10/2017 15:00 | 26/10/2017 16:00 | 26/10/2017 21:46 | 08/11/2017 22:22 | 23/11/2017 00:00 |
| 22 | 203096114ee82d0edff4c1e6bc2b5f | delivered | 18/09/2017 14:31 | 19/09/2017 04:04 | 06/10/2017 17:50 | 09/10/2017 22:23 | 28/09/2017 00:00 |
| 23 | f848643ee16939509eb3840d2051 | delivered | 15/03/2018 08:52 | 15/03/2018 09:09 | 15/03/2018 19:52 | 19/03/2018 18:08 | 29/03/2018 00:00 |
| 24 | 2807d0c504d48944d1672727b139f | delivered | 03/02/2018 20:37 | 03/02/2018 20:50 | 05/02/2018 22:37 | 08/02/2018 16:13 | 21/02/2018 00:00 |
| 25 | 95266dfbf7e20354bab07964dc78d5 | delivered | 08/01/2018 07:55 | 08/01/2018 08:07 | 24/01/2018 23:16 | 26/01/2018 17:32 | 21/02/2018 00:00 |
| 26 | f3e7359154d9658273551396b1f1dac | delivered | 09/08/2018 11:44 | 10/08/2018 03:24 | 10/08/2018 12:29 | 13/08/2018 18:24 | 17/08/2018 00:00 |
| 27 | fb9fc61453ac66c8ea9783d7d0a6f | delivered | 20/02/2018 23:46 | 22/02/2018 02:30 | 26/02/2018 22:25 | 21/03/2018 00:00 | 12/03/2018 00:00 |

1.1.4 Order Payment Dataset

This dataset includes data about the orders payment options.

Figure 5: Order payment dataset viewed in excel

| order_id | payment_sequential | payment_type | payment_installments | payment_value |
|-------------------------------------|--------------------|--------------|----------------------|---------------|
| b81ef26f3fe1789b1e8b2acac839d17 | | credit_card | 8 | 99.33 |
| a9810da82917af2d9aef1278f1dcfa0 | | credit_card | 1 | 24.39 |
| 4 25e8ea4e93396b6fa0d3d708e76c1bd | | credit_card | 1 | 65.71 |
| 5 ba78997921bbcd1373bb41e913ab953 | | credit_card | 8 | 107.78 |
| 6 42ff8080a16b47b59251dd489d4441a | | credit_card | 2 | 128.45 |
| 7 298fcfd1f173eb413ed42d601b25b1cd | | credit_card | 2 | 96.12 |
| 8 71lee386b001f06208a7419e4fc1bbd7 | | credit_card | 1 | 81.16 |
| 9 3d7239c394a212faaa122962df514ac7 | | credit_card | 3 | 51.84 |
| 10 1f78449c87a54fa9e96e88ba1491fa9 | | credit_card | 6 | 341.09 |
| 11 0573b5e23cb79806520e1d5b46714 | | boleto | 1 | 51.95 |
| 12 d88e0d5fa1661c03fcf336527646 | | credit_card | 8 | 188.73 |
| 13 2480f727e869fdeb397244a21b721b67 | | credit_card | 1 | 141.9 |
| 14 616105c9352a9668c88303a44e056cd | | credit_card | 1 | 75.78 |
| 15 cf95215a722f3ebf29e6bbab87a29e61 | | credit_card | 5 | 102.66 |
| 16 769214176682788a92801d8907f1b40 | | credit_card | 4 | 105.28 |
| 17 12e5cfe0e4716b59afbe0f4a3b6570 | | credit_card | 10 | 157.45 |
| 18 61059985a6fc0ad6a95d994caacdad | | credit_card | 1 | 132.04 |
| 19 79da3f5e31a1e454f0695dc032ad5 | | credit_card | 1 | 98.94 |
| 20 8acf09207f415d5sacf02df7dfa895c | | credit_card | 4 | 244.15 |
| 21 b2349a3f20dfbeef62e7b31baa22f84b | | credit_card | 3 | 136.71 |
| 22 5a1f6d2277dfb061e72916b9a687a1 | | credit_card | 4 | 47.69 |
| 23 4214cd5a50ece8ee64a1f459dc33a8c | | credit_card | 2 | 170.57 |
| 24 adfbfc2a620196f9a3e747491c7b27 | | credit_card | 4 | 67.5 |

1.1.5 Order reviews dataset

This dataset includes data about the reviews made by the customers

Figure 6: Order reviews dataset displayed with excel

| review_id | order_id | review_score | review_comment_title | review_comment_message | review_creation_date | review_answer_timestamp |
|-------------------------------------|----------------------------------|--------------|------------------------|---------------------------------------|----------------------|-------------------------|
| 2 7bc2406110b926393aa5f68040eba40 | 73fc7af87114b39712e6da79b0a377eb | 4 | | | 18/01/2018 00:00 | 18/01/2018 21:46 |
| 3 80641a11e56f08060ad469d5645fdfe | m548910a1a6147796b98fd73dbdea33 | 5 | | | 10/03/2018 00:00 | 11/03/2018 03:05 |
| 4 282ce5500d1c8e020d8d1322874bf0 | 9e4b658b1a92f9ecdecbb34bed034b | 5 | | | 17/02/2018 00:00 | 18/02/2018 14:36 |
| 5 e64fb393e7b3283ab78ff8bb30750e | 658677c97b385a9be170737859d3511b | 5 | | Recebi bem antes do prazo estipulado. | 21/04/2017 00:00 | 21/04/2017 22:02 |
| 6 f7d4243cfe1938f181bce41a392bdeb | 8e6b6fb1e283f7e4f11123a3fb894f1 | 5 | | Parabéns lojas lannister adorei co | 01/03/2018 00:00 | 02/03/2018 10:26 |
| 7 15197a66f4d0650b5434f1b46cd19 | b18dcdf73be66366873cd26c5724d1dc | 1 | | | 13/04/2018 00:00 | 16/04/2018 00:39 |
| 8 079bbe5d1b850860ad761afa7ff16 | e48a4od2dce3a2e87348811bcfd22b | 5 | | | 16/07/2017 00:00 | 18/07/2017 19:30 |
| 9 76400515c6769fbbee952a752581ef | c31a859e34e3ada22f376954e19b39d | 5 | | | 14/08/2018 00:00 | 14/08/2018 21:36 |
| 10 a36f7f6433de0aefbb97da197c554 | 9c21ac490e8427358a5b23dfad909b | 5 | | | 17/05/2017 00:00 | 18/05/2017 12:05 |
| 11 8670ds2e150043ae7de40c12ce0f | b9bf720beb4ab372876008858962129 | 4 | recomendo | aparelho eficiente. no site a marca c | 22/05/2018 00:00 | 23/05/2018 16:45 |
| 12 c9cd2d5ab5911836ababae1363a10c | cd9fa9a8e72324eeb25c7de974696ee2 | 5 | | | 23/12/2017 00:00 | 26/12/2017 14:36 |
| 13 96052551d87e5f62e69f6974ec392e9 | 3d374c9e46530bb5ed4a7648915306a6 | 5 | | | 19/12/2017 00:00 | 20/12/2017 10:25 |
| 14 4b49719c8a200003f700d3d980ea119 | 9d6f15f95d01e79b1d349cc208361f09 | 4 | | Mas um pouco...travado...pelo | 16/02/2018 00:00 | 20/02/2018 10:52 |
| 15 23f75a37eff35d9a15b4e1ad483793 | 2eafb9e99d871c5d22b385a8e6fe0e | 4 | | | 28/03/2018 00:00 | 30/03/2018 15:10 |
| 16 90aabbb668abfb9546d2b05db43284c | d7b0de4ad94846eb73d462b4e3e75c3 | 3 | | | 30/04/2017 00:00 | 03/05/2017 00:02 |
| 17 3948b09f7e277a8734b6f991fdbfa3b | 5 Super recomendo | | | Vendedor confiável, produto ok e é | 23/05/2018 00:00 | 24/05/2018 03:00 |
| 18 9314d6f7995fbfa510c7bdc46801 | 0dacf045ad9f5d5a0c1faa7034e39 | 2 | | GOSTARIA DE SABER O QUE HOUVE, | 18/01/2018 00:00 | 20/01/2018 21:25 |
| 19 8e15a72d95600f1a48b64e37a067 | f1f581e083011021e7cf6592ddc81e | 5 | | | 24/03/2018 00:00 | 26/03/2018 15:58 |
| 20 fdb8b297a4d06657acc92084e7f | 70a752414a3d090c1f2a4379b14b28e | 3 | | | 29/09/2017 00:00 | 02/10/2017 01:12 |
| 21 373cbceea8286a2b66c97b1b157e46 | 583174fb373d5d06661be3aa1d1786 | 1 | Não chegou meu produto | PÂOssimo | 15/08/2018 00:00 | 15/08/2018 04:10 |
| 22 a2ff510ca4e059c1efc7a866571f6 | a9953b2f0b1665ac7856cc5a6acbf58c | 4 | | | 06/10/2017 00:00 | 08/10/2017 22:36 |
| 23 80c70b2c63881ba6525f6ceee1d94ac7 | ae168dfba236919411fe08f336fe44 | 5 | | | 17/03/2018 00:00 | 04/04/2018 19:41 |
| 24 d21bbc789670eab777d27372ab904cc | 4f44d78867142c627497b60a7e0228a | 5 | À timo | Loja nota 10 | 10/07/2018 00:00 | 11/07/2018 14:10 |

1.1.6 Order Dataset

This is the core dataset. From each order you might find all other information.

Figure 7: Order dataset viewed in excel

| order_id | customer_id | order_status | order_purchase_timestamp | order_approved_at | order_delivered_carrier_date | order_delivered_customer_date | order_estimated_delivery_date |
|--------------------------------------|-----------------------------------|--------------|--------------------------|-------------------|------------------------------|-------------------------------|-------------------------------|
| e481f51cbcd54678b7cc49136f2d6af7 | 9ef432ebe6251297304e76186b10a928d | delivered | 02/10/2017 10:56 | 02/10/2017 11:07 | 04/10/2017 19:55 | 10/10/2017 21:25 | 18/10/2017 00:00 |
| 53dc262bc8c7dc0b86741e2150273451 | b0830fb47a7a66c620de0b8c802d7ef | delivered | 24/07/2018 20:41 | 26/07/2018 03:24 | 26/07/2018 14:31 | 07/08/2018 15:27 | 13/08/2018 00:00 |
| 47770e09100c2d0c4496d9f07ec5b5d | 41ce2a54c0b03bf3443c3d931a3e7089 | delivered | 08/08/2018 08:38 | 08/08/2018 08:55 | 08/08/2018 13:50 | 17/08/2018 18:06 | 04/09/2018 00:00 |
| 9495b544dbf5ed918fe9c16f9745l8a | 18819746e5e7920adcbec73753d482 | delivered | 18/11/2017 19:28 | 18/11/2017 19:45 | 22/11/2017 13:39 | 02/12/2017 00:28 | 15/12/2017 00:00 |
| a621c590c840e6e8cb3a9cebc573f8159 | aa9797904e6daea88656dbdc4f7baad2c | delivered | 13/02/2018 21:18 | 13/02/2018 22:20 | 16/02/2018 19:46 | 16/02/2018 18:17 | 26/02/2018 00:00 |
| 7e4592615e18b1c1dcee52889e28acc3 | 503730e09ca751ccddab28e9abf608 | delivered | 09/07/2017 21:57 | 09/07/2017 22:10 | 11/07/2017 14:58 | 26/07/2017 10:57 | 01/08/2017 00:00 |
| 8_136cce7fa7af42fd82cef53dc79a6098 | ed0271e0b7da06093796590e7b737a | invoiced | 11/04/2017 12:22 | 13/04/2017 13:25 | | | 09/05/2017 00:00 |
| 9_6514bbad0289f2c2374ed245783f | 9bf0ff8ba3b52b526ff42d374d722 | delivered | 16/05/2017 13:10 | 16/05/2017 13:22 | 22/05/2017 10:07 | 26/05/2017 12:55 | 07/06/2017 00:00 |
| 10_76866289321a7931b548526d33 | f54a90e6b0351431402b8461aa5199 | delivered | 23/01/2017 18:29 | 25/01/2017 02:50 | 26/01/2017 14:16 | 02/02/2017 14:08 | 06/03/2017 00:00 |
| 11_e9fbfb8e88e0edfa78558527e16dfb | 31ad1d1b63eb9962463764de6e0c9d | delivered | 29/07/2017 11:55 | 29/07/2017 12:00 | 10/08/2017 19:45 | 16/08/2017 17:14 | 23/08/2017 00:00 |
| 12_efc616cb79e9c164e05b573f8159 | 494dded5b20131c4ed7f00595b95c | delivered | 16/05/2017 19:41 | 16/05/2017 19:50 | 18/05/2017 11:18 | 29/05/2017 11:18 | 07/06/2017 00:00 |
| 13_34513ce04fa4b62a55830c08987c6b | 7711c6f24183a843aafe81855097bc37 | delivered | 13/07/2017 19:58 | 13/07/2017 20:10 | 14/07/2017 18:43 | 19/07/2017 14:04 | 08/08/2017 00:00 |
| 14_8256a660a92b15fb8ee004a832918 | d3e3b74a76b6214a0c830b1.7ee2341 | delivered | 07/06/2018 10:06 | 09/06/2018 03:13 | 11/06/2018 13:29 | 19/06/2018 12:05 | 18/07/2018 00:00 |
| 15_5f0fc55d0b717a16c5e1d3f477225a250 | 19402a4e0860416a0f932a8e2a27740 | delivered | 25/07/2018 17:44 | 25/07/2018 17:55 | 26/07/2018 13:16 | 30/07/2018 15:52 | 08/08/2018 00:00 |
| 16_432fa71d851672c86e9448c4e42c | 3df70f513d1f144818480d34e672a9f | delivered | 01/03/2018 14:14 | 01/03/2018 15:10 | 02/03/2018 21:09 | 12/03/2018 23:36 | 21/03/2018 00:00 |
| 17_deb3b6511fae05f0697-d5c05de4b2c | 3b6828a50f5e464027a747d70a0cfc | delivered | 07/06/2018 19:03 | 12/06/2018 23:31 | 11/06/2018 14:54 | 21/06/2018 15:34 | 04/07/2018 00:00 |
| 18_402b97836b0c04622354c5f31062e5f | 738b086814c6fc74b8cc583f8515ee3 | delivered | 02/01/2018 19:00 | 02/01/2018 19:09 | 03/01/2018 18:19 | 20/01/2018 01:38 | 06/02/2018 00:00 |
| 19_116fb99343b6d34e6d82f1e29044403 | 31877899e699987628d7a9b0e4dd6ac | delivered | 26/12/2017 23:41 | 26/12/2017 23:50 | 28/12/2017 18:33 | 08/01/2018 22:36 | 29/01/2018 00:00 |
| 20_89c5859f6d63d4e6d82f1e29044403 | 0597fc5179c1d6abcbafe370971a8d70 | delivered | 21/11/2017 00:03 | 21/11/2017 00:14 | 23/11/2017 21:32 | 27/11/2017 18:28 | 11/12/2017 00:00 |
| 21_83018e8714ee8b7a7e0877d4e926f | f7b809c2ae27f0300f670c3d47f288e | delivered | 26/10/2017 15:54 | 26/10/2017 16:08 | 26/10/2017 21:46 | 08/11/2017 22:22 | 23/11/2017 00:00 |
| 22_269969030382ed0ff7a1e6b3b6fe | d2b091513d1f24a1b3a12c18bc3b6fe | delivered | 18/09/2017 14:31 | 19/09/2017 04:04 | 06/10/2017 17:50 | 09/10/2017 22:23 | 28/09/2017 00:00 |
| 23_1848643ee1693959e5384002051 | 4f41cd16a598b6de80f48aae4e43 | delivered | 15/03/2018 08:52 | 15/03/2018 09:09 | 15/03/2018 19:52 | 19/03/2018 18:08 | 29/03/2018 00:00 |
| 24_2807d0504d6489441672727bc139f | 72ae281627a6102d99371852b40208a | delivered | 03/02/2018 20:37 | 03/02/2018 20:50 | 05/02/2018 22:37 | 08/02/2018 16:13 | 21/02/2018 00:00 |
| 25_9526d60b9a7b15fb8ee004a832918 | 1a06d3a4890079a942054b3e426 | delivered | 08/01/2018 07:55 | 08/01/2018 08:07 | 24/01/2018 23:16 | 26/01/2018 17:32 | 21/02/2018 00:00 |
| 26_f3e739154a965827355193d6b1f1dc | 626423aa5b8096c5a614ba6aa6b6e2f98 | delivered | 09/08/2018 11:44 | 10/08/2018 03:24 | 10/08/2018 12:29 | 13/08/2018 18:24 | 17/08/2018 00:00 |
| 27_fb9a6c1453ac64cc8a9d783d7d0a6f | 3a3b74a4d4c6b543206f5d892870c3 | delivered | 20/02/2018 23:46 | 22/02/2018 02:30 | 26/02/2018 22:25 | 21/03/2018 00:00 | 21/03/2018 00:00 |

1.1.7 Products Dataset

This dataset contains data about the products sold by Olist.

Figure 8: Products dataset viewed in excel

| product_id | product_category_name | product_name_length | product_description_length | product_photos_qty | product_weight_g | product_length_cm | product_height_cm | product_width_cm |
|---------------------------------------|------------------------|---------------------|----------------------------|--------------------|------------------|-------------------|-------------------|------------------|
| 2_1e8e0404dbcff541ed26657ea517e5 | perfumaria | 40 | 287 | 1 | 225 | 16 | 10 | 14 |
| 3_aa071139b1b6b7aca5deaa41aa2f | artes | 44 | 276 | 1 | 1000 | 30 | 18 | 20 |
| 4_96b6ee8810374e1b65e391975717f | esporte_lazer | 46 | 250 | 1 | 154 | 18 | 9 | 15 |
| 5_ce6f7bcf19066a932b763e39eb23d | bebés | 27 | 261 | 1 | 371 | 26 | 4 | 26 |
| 6_9dc1a7de27444849c19cff195d0b71 | utilidades_domesticas | 37 | 402 | 4 | 625 | 20 | 17 | 13 |
| 7_41d3672d4792049fa179bb35283ed13 | instrumentos_musicais | 60 | 745 | 1 | 200 | 38 | 5 | 11 |
| 8_732bd381ad09530fe0d457d81becc | cool_stuff | 56 | 1272 | 4 | 18350 | 70 | 24 | 44 |
| 9_2548a13e677a690f3e6b368e9ab61 | moveis_decoracao | 56 | 184 | 2 | 900 | 40 | 8 | 40 |
| 10_37c742b0e7708853a870267a1d0 | eletronicos | 57 | 163 | 1 | 400 | 27 | 13 | 17 |
| 11_892109888e8df9d66cf463025574 | brinquedos | 36 | 1156 | 1 | 600 | 17 | 10 | 12 |
| 12_14a47b7fe525522b47b74b29c98dc9b | cama_mesa_banho | 54 | 630 | 1 | 1100 | 16 | 10 | 16 |
| 13_03b63c5f16691530586a020c345514 | bebés | 49 | 728 | 4 | 7150 | 50 | 19 | 45 |
| 14_cf5509e8adeaaa1d28fb16e48fc2 | Instrumentos_musicais | 43 | 1827 | 3 | 250 | 17 | 7 | 17 |
| 15_7bb6f79cbe5771619496496607c2 | moveis_decoracao | 51 | 2083 | 2 | 600 | 68 | 11 | 13 |
| 16_eb3143580a610f202c8594638c7415 | construcao_ferramentas | 59 | 1602 | 4 | 200 | 17 | 7 | 17 |
| 17_3bb7f144026e673277d8d838a7b13b3 | esporte_lazer | 22 | 3021 | 1 | 800 | 16 | 2 | 11 |
| 18_62fb4dd53d2cd88e0432f1284a004c | perfumaria | 39 | 346 | 2 | 400 | 27 | 5 | 20 |
| 19_ab17017a495f2d8da4eeefba108a24 | informatica_acessorios | 59 | 636 | 1 | 900 | 40 | 15 | 20 |
| 20_a073692e326eecd290e30b578413b6 | moveis_decoracao | 56 | 296 | 2 | 1700 | 100 | 7 | 15 |
| 21_f53103a7d79c2456579ea37e5c51f0 | cama_mesa_banho | 52 | 206 | 1 | 500 | 16 | 10 | 16 |
| 22_1c1890ba1779090cd54008a3c302921 | moveis_decoracao | 27 | 158 | 4 | 2550 | 29 | 24 | 45 |
| 23_518ef5dec2b3a553e26a4594ba15d9 | cama_mesa_banho | 27 | 329 | 2 | 800 | 36 | 8 | 16 |
| 24_e3e020a31d314d488d2602272b315c3f6e | beleza_saude | 58 | 1987 | 3 | 75 | 21 | 7 | 13 |

1.1.8 Sellers Dataset

This dataset contains data about the sellers that fulfilled orders made at Olist.

Figure 9: Sellers' dataset viewed in excel

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|----------------------------------|-----------------|--------|-------------------------|--------------|---|---|---|---|---|---|---|---|---|
| 1 | seller_id | seller_zip_code | prefix | seller_city | seller_state | | | | | | | | | |
| 2 | 34428959a84dea7ee197c632cb2df15 | | | 13023 campinas | SP | | | | | | | | | |
| 3 | d1b65fc7deb3361ea86b5f14c68d2e2 | | | 13844 mogi guacu | SP | | | | | | | | | |
| 4 | ce3ad9de960102d0677a81fbdb7b2d | | | 20031 rio de janeiro | RJ | | | | | | | | | |
| 5 | c0f3ee2e21455b6faeae3dd58c1b1c3 | | | 4195 sao paulo | SP | | | | | | | | | |
| 6 | 51a04a8a6bdc2b3dec282b0b80742cf | | | 12914 braganca paulista | SP | | | | | | | | | |
| 7 | c240c4061717aca1806a6ee72be3533b | | | 20290 rio de janeiro | RJ | | | | | | | | | |
| 8 | e49c263cedfa6d227d5121a6b6e4d37 | | | 55325 brejao | PE | | | | | | | | | |
| 9 | 1b938a7ec6a5061a66a3766e0e75f90 | | | 16304 penapolis | SP | | | | | | | | | |
| 10 | 768a86e36ad6aae3d03ee3c6433d61f | | | 1529 sao paulo | SP | | | | | | | | | |
| 11 | cc4bb5bf32a6ab27066a4130f114e3 | | | 80310 curitiba | PR | | | | | | | | | |
| 12 | 8cb7c5df4f14d506eba769a4702a25 | | | 75110 anapolis | GO | | | | | | | | | |
| 13 | a7a9b880c49781da66651cc4ba9a38 | | | 13530 itrapina | SP | | | | | | | | | |
| 14 | 8bd0f31c0a614c658f6763bd02dea69 | | | 1222 sao paulo | SP | | | | | | | | | |
| 15 | 05a48cc8859962767935ab9087417bb | | | 5372 sao paulo | SP | | | | | | | | | |
| 16 | 7bb8e8c35bad4b0e7e3963650bba87b | | | 88705 tubarao | SC | | | | | | | | | |
| 17 | 1444c08e64d55b3c25f0f09e7ffcf2 | | | 42738 lauro de freitas | BA | | | | | | | | | |
| 18 | 166e8f1381e0951983c38b1f691c11 | | | 88780 imbituba | SC | | | | | | | | | |
| 19 | e38db885400cd35c1df162f2c1dbc | | | 70740 brasilia | DF | | | | | | | | | |
| 20 | d2e753b80b7d4faa77483ed0edc8a | | | 45810 porto seguro | BA | | | | | | | | | |
| 21 | f9ec7093df3a7b346b7cfc7864069ca3 | | | 5138 sao paulo | SP | | | | | | | | | |
| 22 | 46e015589b781ada5ce1f892d06b1 | | | 11440 guaruja | SP | | | | | | | | | |
| 23 | 4cf90a58259286ada5ba8525ba9e84a | | | 14910 tabatinga | SP | | | | | | | | | |
| 24 | f7496659ca9fdf323da0ae84176632 | | | 4156 sao paulo | SP | | | | | | | | | |

1.1.9 Category Name Translation

Translates the “productcategoryname” to English.

Figure 10: Category name translation dataset viewed in excel

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|-----------------------------|-------------------------------|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | product_category_name | product_category_name_english | | | | | | | | | | | | |
| 2 | beleza_saude | health_beauty | | | | | | | | | | | | |
| 3 | informatica_acessorios | computers_accessories | | | | | | | | | | | | |
| 4 | automotivo | auto | | | | | | | | | | | | |
| 5 | cama_mesa_banho | bed_bath_table | | | | | | | | | | | | |
| 6 | moveis_decoracao | furniture_decor | | | | | | | | | | | | |
| 7 | esporte_lazer | sports_leisure | | | | | | | | | | | | |
| 8 | perfumaria | perfumery | | | | | | | | | | | | |
| 9 | utilidades_domesticas | housewares | | | | | | | | | | | | |
| 10 | telefonia | telephony | | | | | | | | | | | | |
| 11 | relogios_presentes | watches_gifts | | | | | | | | | | | | |
| 12 | alimentos_bebidas | food_drink | | | | | | | | | | | | |
| 13 | bebés | baby | | | | | | | | | | | | |
| 14 | papelaria | stationery | | | | | | | | | | | | |
| 15 | tablets_impressao_imagem | tablets_printing_image | | | | | | | | | | | | |
| 16 | brinquedos | toys | | | | | | | | | | | | |
| 17 | telefonia_fixa | fixed_telephony | | | | | | | | | | | | |
| 18 | ferramentas_jardim | garden_tools | | | | | | | | | | | | |
| 19 | fashion_bolsas_e_acessorios | fashion_bags_accessories | | | | | | | | | | | | |
| 20 | eletroportateis | small_appliances | | | | | | | | | | | | |
| 21 | consoles_games | consoles_games | | | | | | | | | | | | |
| 22 | audio | audio | | | | | | | | | | | | |
| 23 | fashion_calcados | fashion_shoes | | | | | | | | | | | | |
| 24 | cool_stuff | cool_stuff | | | | | | | | | | | | |

1.2 Brazilian States Information Dataset

The location information available in the olist dataset is limited, and there is need to explore more dimensions of the available

The Brazil dataset is sourced from Brazil-Help.com. The website was created to provide English-speaking people around the world with information, knowledge, assistance, help, and practical tips about Brazil and its people.

https://brazil-help.com/brazilian_states.htm

Figure 11: Dataset containing Brazilian States Information as seen on the website

Bom dia! Hoje é terça-feira, 28 de dezembro de 2021 • Good morning! Today is Tuesday, December 28, 2021

For political administration, Brazil's 26 states—and the Federal District (Distrito Federal, Brasília)—are divided into [five different regions](#). The name of each of the Brazilian states below is preceded by its commonly used two-letter abbreviation. The chart also includes the state's capital city, its size in km² (square kilometers), 2007 population (the latest available), percentage of urban and rural population, number of municipal districts (cities), per capita GNP and projected life expectancy (the latest available). For the specific location of each individual state within Brazil, see the [state's regions](#) or see the [political map of Brazil](#).

Kia [Kia](#) [Kia](#)

DISCOUNT CODE AVAILABLE [Continue →](#)

Click the "Correios" logo to look up Brazilian CEPs (zip codes).

CORREIOS [Telephone Numbers](#) look up telephone numbers in Brazil

| BR | Brazil (entire country) | Brasília | - | 8,514,204 (slightly larger than the Continental United States) | 192,000,000 (2008 estimate) | 82.5 / 17.5 | 5,564 | R\$8,300 | 72.24 |
|--------------------------------|-------------------------|--------------|---------------------------|--|-----------------------------|--------------------|-------------------------------|-------------------------------|-----------------------------------|
| Common Two Letter Abbreviation | State | Capitol City | Region | Size (in km ²) | Population (2007 estimate) | % Pop. Urban/Rural | Number of Municipal Districts | Per Capita GNP in Reais (R\$) | Life Expectancy (2007 projection) |
| AC | Acre | Rio Branco | North | 152,581 | 664,000 | 69.6 / 30.4 | 24 | R\$5,413 | 71.1 |
| AL | Alagoas | Maceló | Northeast | 27,767 | 3,557,000 | 67.4 / 32.6 | 102 | R\$3,876 | 65.4 |

1.3 Summary of Dataset

Table 1: Showing description of all the tables that make up the Olist dataset

| S N | Table Title | No. of rows | No. of columns | Columns Title | Description | Include? |
|-----|-------------------------|-------------|----------------|------------------------------|-------------------------------------|----------|
| 1 | Olist_customers_dataset | 99441 | 5 | Customer_city | City customer is located | Yes |
| | | | | Customer_id | Customer ID number | Yes |
| | | | | Customer_state | State customer is located | Yes |
| | | | | Customer_unique_id | Customer's unique ID number | Yes |
| | | | | Customer_zip_code_prefix | First 4 digits of customer zip code | Yes |
| 2 | olist_orders_data set | 99441 | 8 | Customer_id | Customer ID number | Yes |
| | | | | Order_approved_at | Time of order approval | Yes |
| | | | | Order_delivered_carrier_date | Date of order pickup by carrier | Yes |

| | | | | | | | |
|---|-----------------------------------|---------|---|--|-------------------------------|---|-----|
| | | | | | Order_delivered_customer_date | Date of order delivery to customer | Yes |
| | | | | | order_estimated_delivery_date | Estimated delivery date | Yes |
| | | | | | order_id | Order ID number | Yes |
| | | | | | order_status | Order delivery status | Yes |
| 3 | olist_order_items_dataset | 112650 | 7 | | freight_value | Cost of delivery | Yes |
| | | | | | order_id | Order ID number | Yes |
| | | | | | order_item_id | Order item ID number | Yes |
| | | | | | price | Price of order item | Yes |
| | | | | | product_id | Product ID number | Yes |
| | | | | | seller_id | Seller ID number | Yes |
| | | | | | shipping_limit_date | Shipping limit date | Yes |
| 4 | olist_order_reviews_dataset | 99224 | 7 | | order_id | Order ID number | Yes |
| | | | | | review_answer_timestamp | Time of review answer | Yes |
| | | | | | review_comment_title | Title of comment | Yes |
| | | | | | review_creation_date | Review start date | Yes |
| | | | | | review_id | Review ID number | Yes |
| | | | | | review_score | Review score | Yes |
| 5 | olist_order_payments_dataset | 103886 | 5 | | order_id | Order ID number | Yes |
| | | | | | payment_installments | Number of instalments | Yes |
| | | | | | payment_sequential | Number of payment methods that a customer may pay | Yes |
| | | | | | payment_type | Payment type | Yes |
| | | | | | payment_value | Payment value | Yes |
| 6 | olist_geolocation_dataset | 1000163 | 5 | | geolocation_city | City | Yes |
| | | | | | geolocation_lat | Latitude | Yes |
| | | | | | geolocation_lng | Longitude | Yes |
| | | | | | geolocation_state | State | Yes |
| | | | | | geolocation_zip_code_prefix | Zip code | Yes |
| 7 | product_category_name_translation | 71 | 2 | | product_category_name | Product category name in Brazil | Yes |
| | | | | | product_category_name_english | Product category name in English | Yes |

| | | | | | | | |
|-----------|--|-------------------------|-------|----|---------------------------------|---|-----|
| 8 | | olist_seller_datas et | 3095 | 4 | seller_city | City the seller is located | Yes |
| | | | | | seller_id | Seller ID number | Yes |
| | | | | | seller_state | State seller is located | Yes |
| | | | | | seller_zip_code_prefix | seller zip code | Yes |
| 9 | | olist_products_d ataset | 32951 | 9 | product_category_nam e | Product category name | Yes |
| | | | | | product_description_le nght | Letter count of product description | Yes |
| | | | | | product_height_cm | Height of product in centimetres | Yes |
| | | | | | product_id | Product ID number | Yes |
| | | | | | product_lenght_cm | Length of product in centimetres | Yes |
| | | | | | product_name_lenght | Letter count of product name | Yes |
| | | | | | product_photos_qty | Number of product photos displayed | Yes |
| | | | | | product_weight_g | Weight of product in grams | Yes |
| | | | | | product_width_cm | Product width in centimetres | Yes |
| 10 | | brazil_info | 27 | 10 | %Pop.Urban/Rural | Ratio of urban to rural population | Yes |
| | | | | | CapitolCity | Capital city | Yes |
| | | | | | CommonTwoLetterAbb reviation | Abbreviation | Yes |
| | | | | | LifeExpectancy(2007 projection) | Life expectancy as of 2007 | No |
| | | | | | Number of Municipal District | Number of municipal districts | Yes |
| | | | | | PerCapita GNP in Reais (R\$) | Economic strength in terms of production value per individual | Yes |
| | | | | | Population (2007 estimate) | Population | Yes |
| | | | | | Region | Region | Yes |
| | | | | | Size (in km-square) | Size in square kilometers | Yes |
| | | | | | State | State | Yes |

1.4 Key Performance Indicators

Based on the Olist Business Model, the platform generates revenue from traffic and sales, which increases with the number of orders made, the number of buyers and sellers, and how much the service is liked by the users. So, the Key Performance Indicators (KPIs) can be measured by these metrics provided the data is available.

With the information present in the Olist dataset, we can categorise the KPIs for this business model as shown below.

1. Sales revenue
2. Number of orders
3. Number of customers
4. Customer satisfaction (Ratings)

1.5 Business Questions

The first objective of this design is to provide access to summarized information on Olist business performance based on measured metrics (KPIs) that have been derived from the available Olist datasets.

Secondly, analysis of these measures with regards to Product Category and Location in order to reveal hidden performance drivers by trying to answer these business questions.

1. What is the summary and trend of Olist KPIs?
2. Where are the main customers from?
3. How does relative distance between sellers and buyers affect KPIs?
4. Does the development and prosperity of a region affect orders?
5. Does product volume affect KPI?
6. Does the number of pictures on a listing affect orders?
7. Which category generates the most revenue?
8. Which categories cause more/less satisfaction?

1.6 User group for this report

The main audience for this report is staff/management of Olist involved in decision-making process to drive business growth. This user group can be categorised into teams.

- Olist Management Team
- Sales and Marketing Team
- Business Development Team

1.7 Why is this information needed?

This report will help the above listed teams to collaboratively create informed sales and marketing strategy with focus on products and location attributes of the available information.

Users of this product will discover key drivers that can be capitalized on for growth plans such as

- Geographical expansion plans
- Incentives generation for customers and sellers
- Development of marketing strategies and user retention programs
- General process monitoring

2 DATA PRE-PROCESSING AND DATA CLEANSING

To create an efficient Business Intelligence Model in Power BI, best practices require well-structured dataset with relationships among datasets with no errors.

To achieve this, we perform some pre-processing steps which involves importing datasets into Power Query, cleaning the data to get rid of unnecessary data and finally building relationships within the dataset for the business intelligence model.

2.1 Importing Datasets

2.1.1 Import Olist Datasets from Excel into Power Query

Use the ‘Get data’ icon on the home ribbon in Power BI and select ‘Excel workbook’ as the data source to import the already-downloaded dataset. Import all 9 tables of the original dataset one after the other into ‘power query’ to transform data and ensure every table has a named header.

Figure 12: Get Data from Excel Workbook

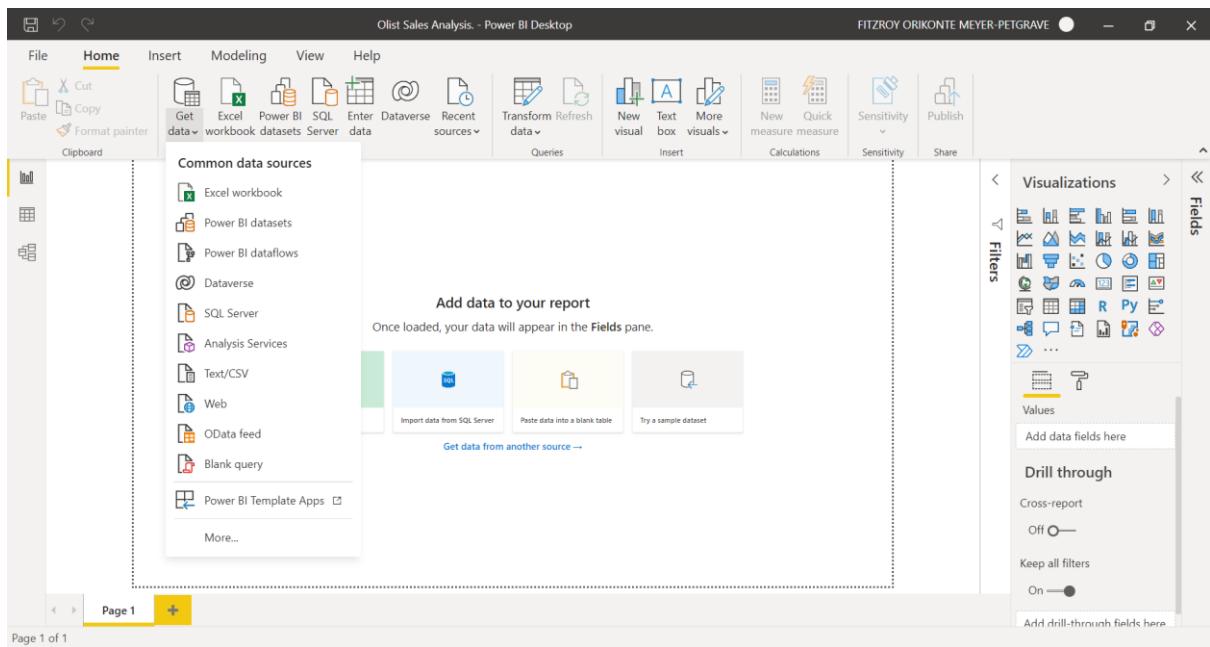


Figure 13: Import files one after the other

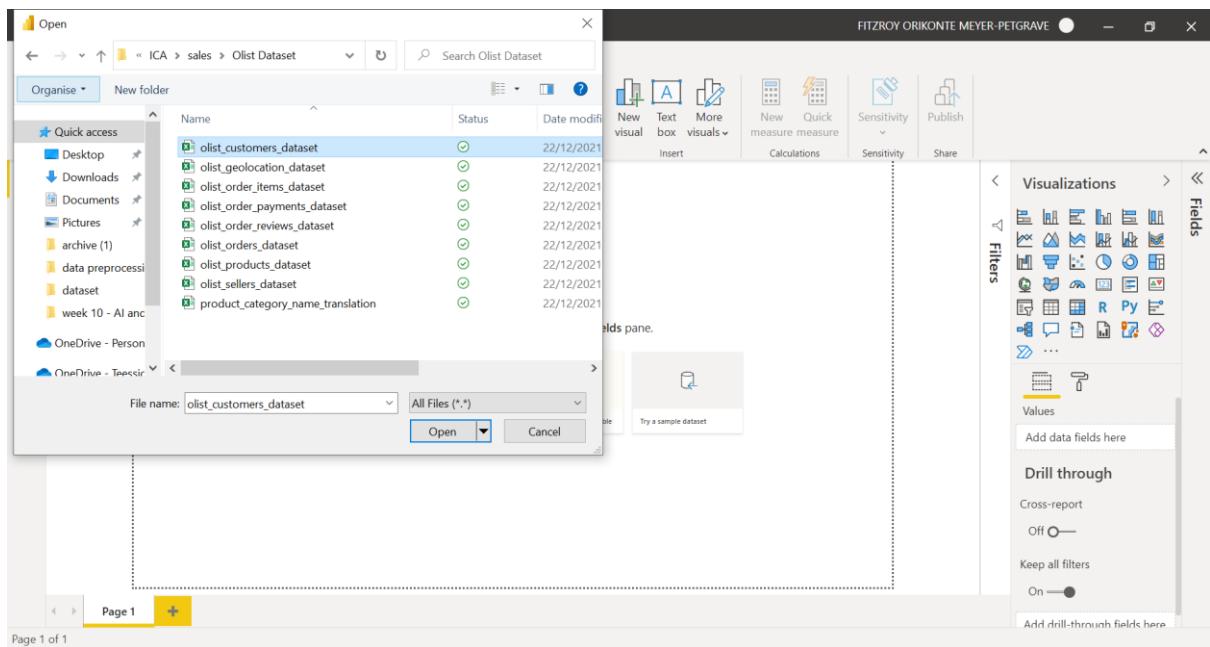
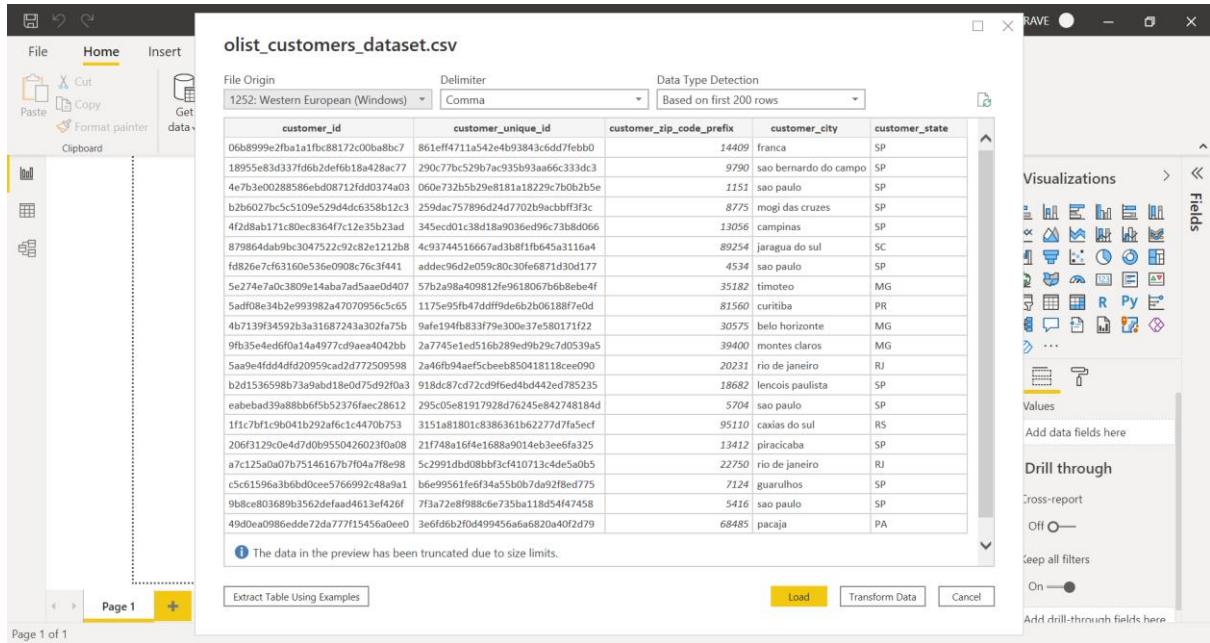


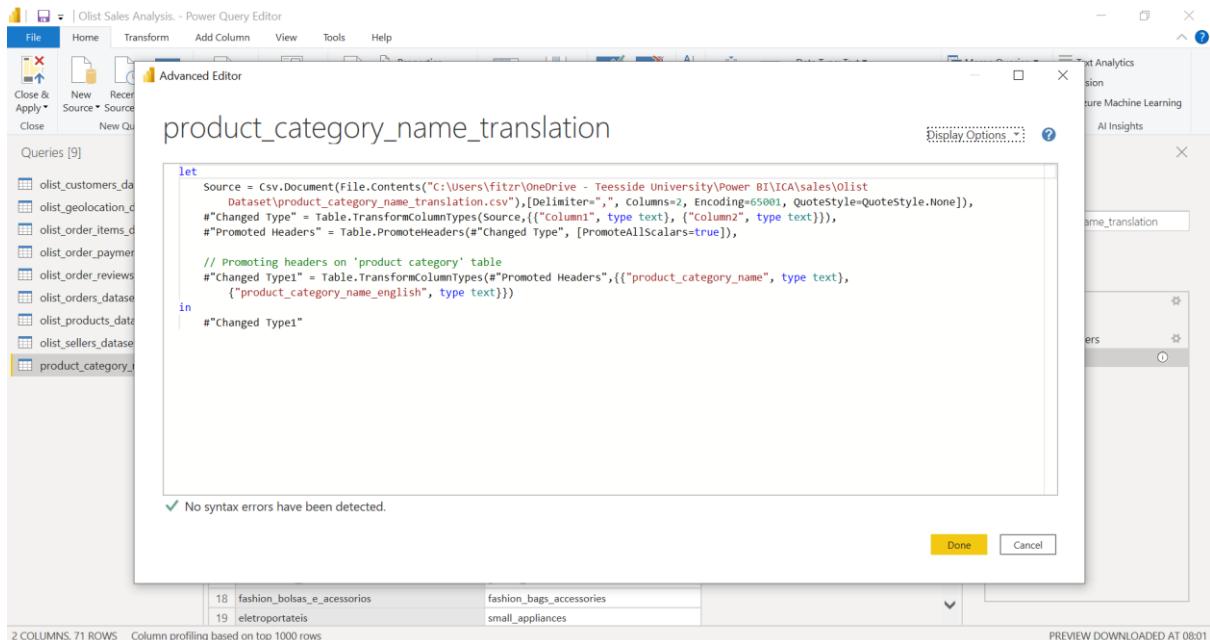
Figure 14: Using 'Transform Data' to open Power Query editor



2.1.2 Use First Rows as Headers

When ‘product names translation’ table was opened its headers were on the 2nd row of the table. We use the ‘first row as headers’ tab to promote the first row to headers. This step is done twice to get the 2nd rows as headers. Below is a snapshot of the M Language for promoting first rows as headers.

Figure 15: M Language for using first rows as headers



2.1.3 Import Brazil Dataset into Power Query from the Web

This dataset is imported into Power BI through Get data function in the Power Query environment. The steps below will import the Brazilian states dataset from the web directly into power query for processing.

- From the ‘get data’ drop-down menu, select ‘from web’ and type in the website address as shown in the snapshot below. Hit ‘ok’.
- In the ‘Navigator’ window that opens, check only table 4, as we do not need any other table from that website.

Figure 16: Load Dataset from the www.brazil-help.com

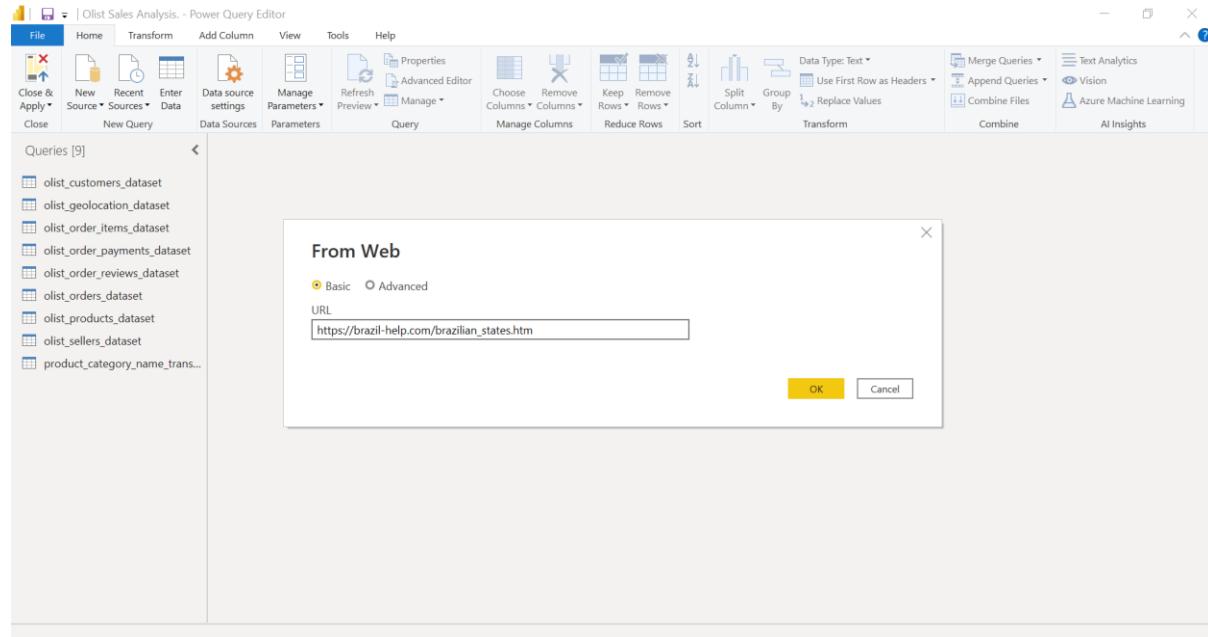


Figure 17: Select required table from the html table folder

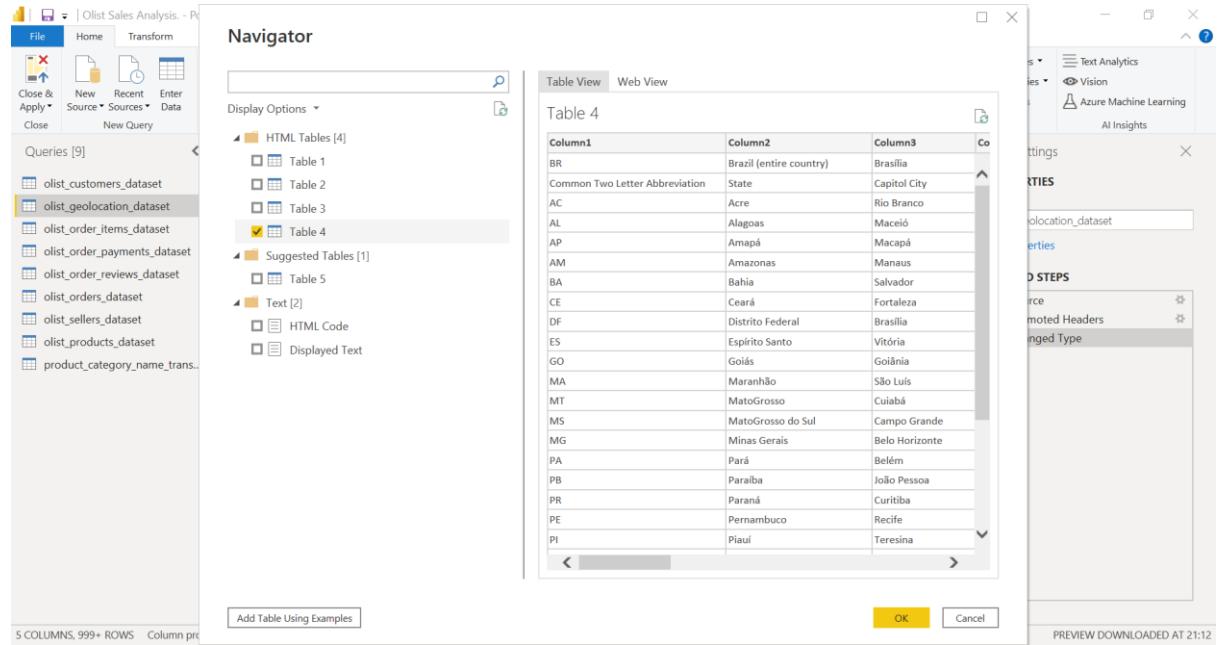


Figure 18: Brazilian states dataset imported into power query from a website

The screenshot shows the Power Query Editor interface. At the top, there are tabs for 'File', 'Home', 'Transform', 'Add Column', 'View', 'Tools', and 'Help'. Below the tabs are buttons for 'Close & Apply', 'New', 'Recent', 'Sources', 'Data', 'Data source settings', 'Manage Parameters', 'Refresh Preview', 'Advanced Editor', 'Properties', 'Choose Columns', 'Remove Columns', 'Keep Rows', 'Remove Rows', 'Sort', 'Split Column', 'Group By', 'Replace Values', 'Data Type: Text', 'Use First Row as Headers', 'Merge Queries', 'Append Queries', 'Combine Files', and 'Combine'. The main area shows a 'Queries [10]' list with items: 'olist_customers_dataset', 'olist_geolocation_dataset', 'olist_order_items_dataset', 'olist_order_payments_dataset', 'olist_order_reviews_dataset', 'olist_orders_dataset', 'olist_sellers_dataset', 'olist_products_dataset', 'product_category_name_trans...', and 'brazil_info'. The 'brazil_info' query is selected and expanded, showing its schema. The schema table has columns: 'Common Two Letter...', 'State', 'Capital City', 'Region', 'Size (in km²)', and 'Population'. The data table below contains 13 rows of state information. At the bottom, it says '10 COLUMNS, 27 ROWS' and 'Column profiling based on top 1000 rows'.

2.2 Data Cleaning and Preparation

In order to have a good model performance, it is necessary to prepare our dataset.

These steps are aimed at exploring functions that are used to combine multiple tables into one large table to create a new model.

2.2.1 Delete Existing Relationships

Figure 19: Deleting existing relationships from initial model

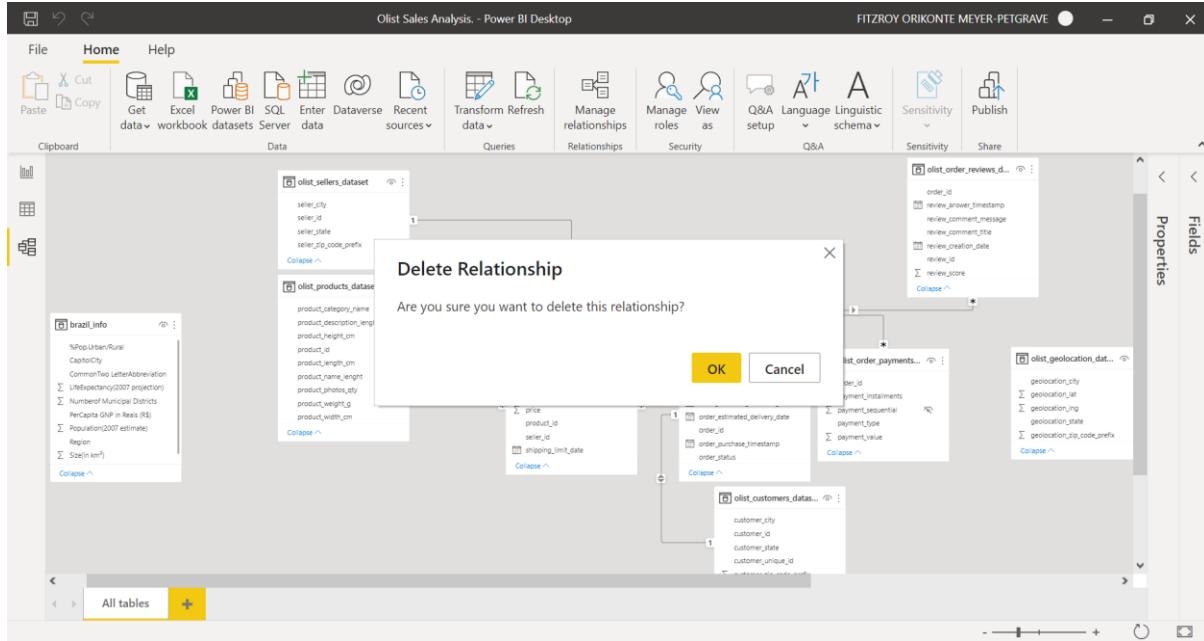
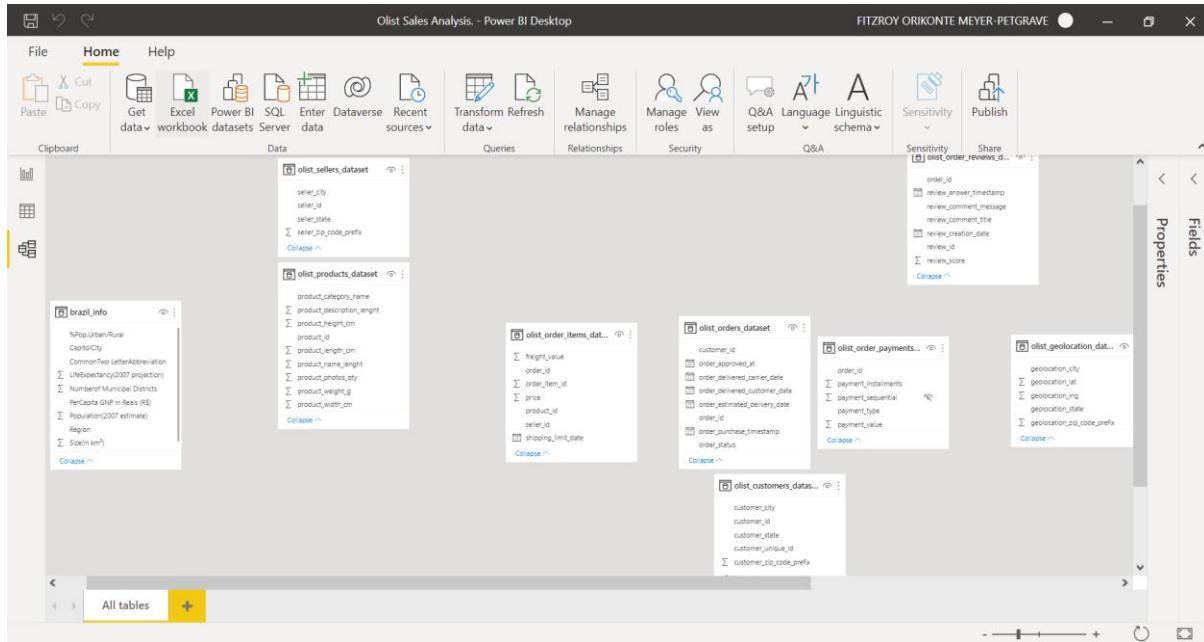


Figure 20: Model showing unlinked tables



2.2.2 Remove Null Values

Empty or invalid cells are irrelevant to our model. It is necessary to identify columns with null values and remove them. To do this we open the ‘view’ tab and check ‘column quality’ and ‘column distribution’. We either remove the rows with invalid cells or

delete the entire columns.

Figure 21: Displaying column quality and distribution to identify null values

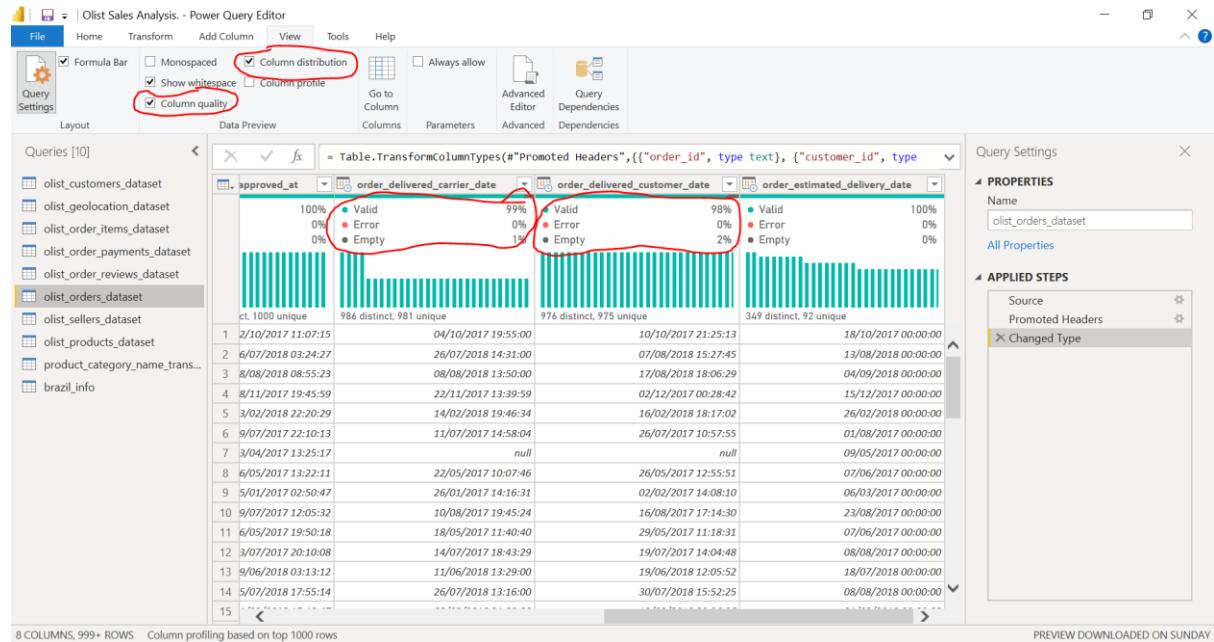


Figure 22: Uncheck null in the drop-down menu to remove Null Values in the column

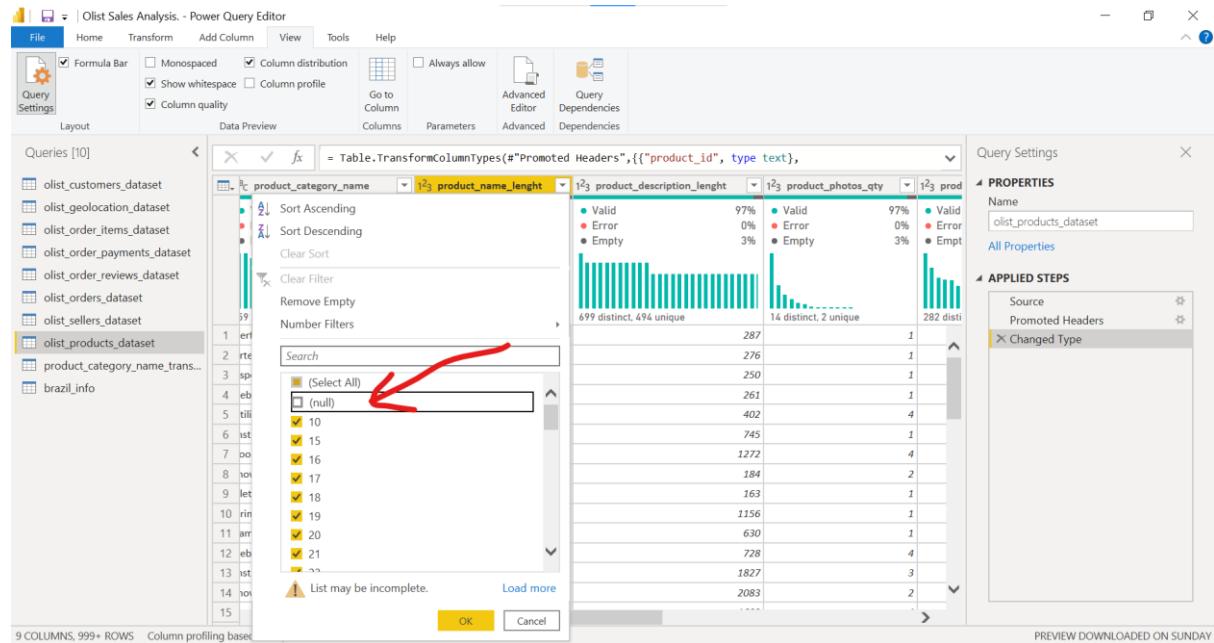
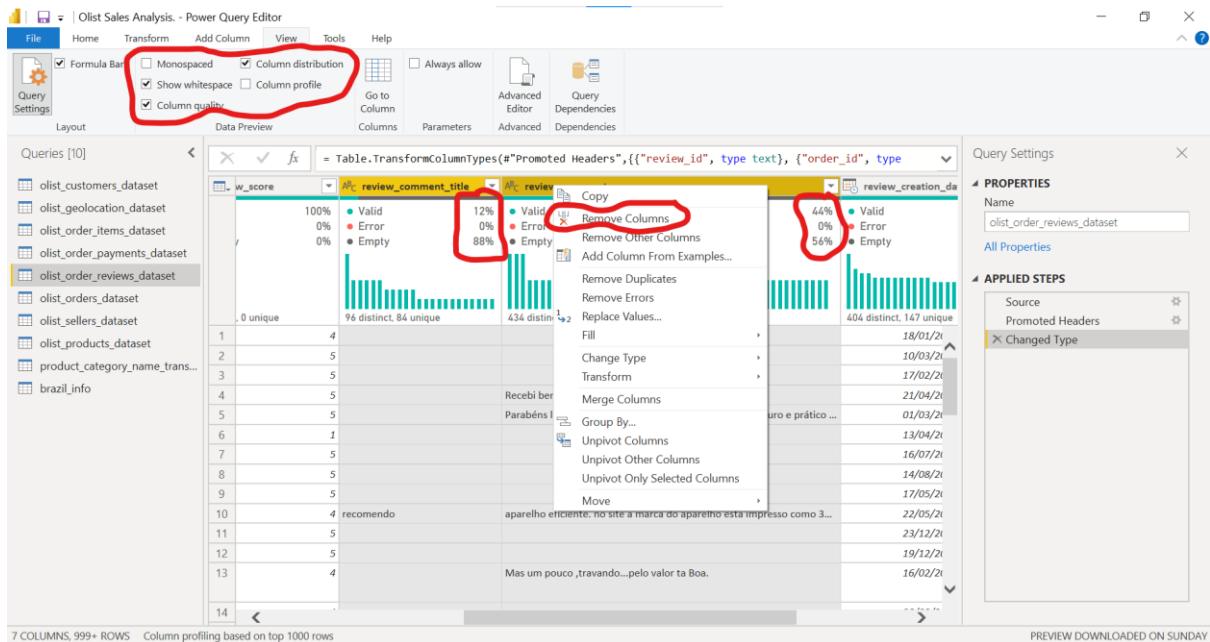
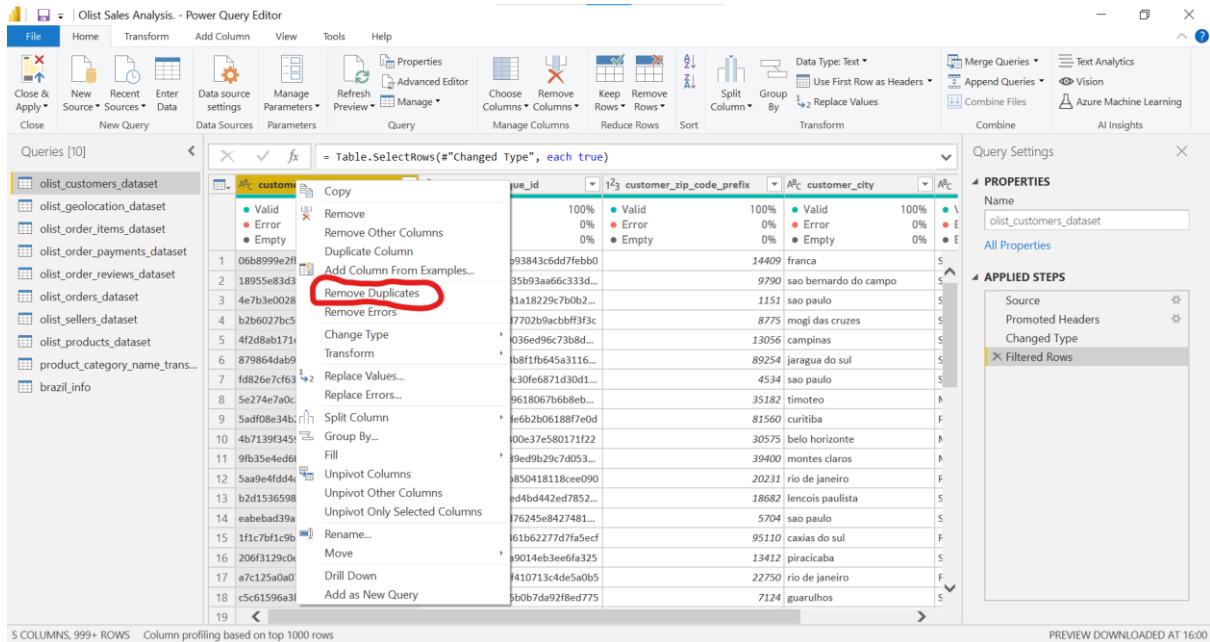


Figure 23: Deleting columns with many empty/invalid cells



2.2.3 Remove Duplicates

Figure 24: Remove duplicates



2.2.4 Join tables using the 'Merge Queries' function

The product dataset contains names of the product categories in Brazilian language. However, there is another dataset with the English translations. So, we join these two tables to see all product information in a single table.

Figure 25: Using the 'Merge Queries' function to join tables by 'product_category_name'

The screenshot shows the Power Query Editor interface. In the ribbon, the 'Merge Queries' button is highlighted with a red circle. The main area displays a 'Merge' dialog with two tables selected: 'olist_products_dataset' and 'product_category_name_translation'. The 'Join Kind' dropdown is set to 'Left Outer (all from first, matching from second)'. The preview pane at the bottom shows the resulting merged table.

Figure 26: Expand new column to display hidden table

The screenshot shows the Power Query Editor interface. In the ribbon, the 'Expand' button is highlighted with a red circle. The main area displays an 'Expand' dialog for a column named 'product_category_name_translations'. The 'Expand' dropdown is set to '(Select All Columns)' and the 'product_category_name_english' checkbox is checked. The preview pane at the bottom shows the expanded table.

2.2.5 Delete Columns

The dataset contains some columns that would be unused in our model. To make the model lighter, it is necessary to delete them using the delete function. The table below shows columns to be deleted and their respective tables titles.

Figure 27: Remove columns by selecting and using the 'remove column' function as shown

2.2.6 Replace Values

Brazil info dataset contain GNP per capita column which is expressed in R\$. But for better analysis it is necessary to express that column as number rather than text. We follow the step below to use the 'replace value' function to remove the prefix 'R\$' from the actual number.

- Select column by clicking on the header and click on the 'replace value' button as shown
- In the pop-up window, type in 'R\$' in 'value to find' cell and leave 'replace with' empty to replace with nothing.

Figure 28: Using the replace value function to remove R\$ sign

2.2.7 Change Data Type

Per Capita GNP in Reais (R\$) is expressed text with the 'ABC' denoted at the header. To change this to whole number, we follow these steps below.

- Click on 'ABC' at the header of the column
- Select 'whole number' as shown in the figure below

Figure 29: Change Data Type to express column as whole number

2.2.8 Split columns by Delimiter

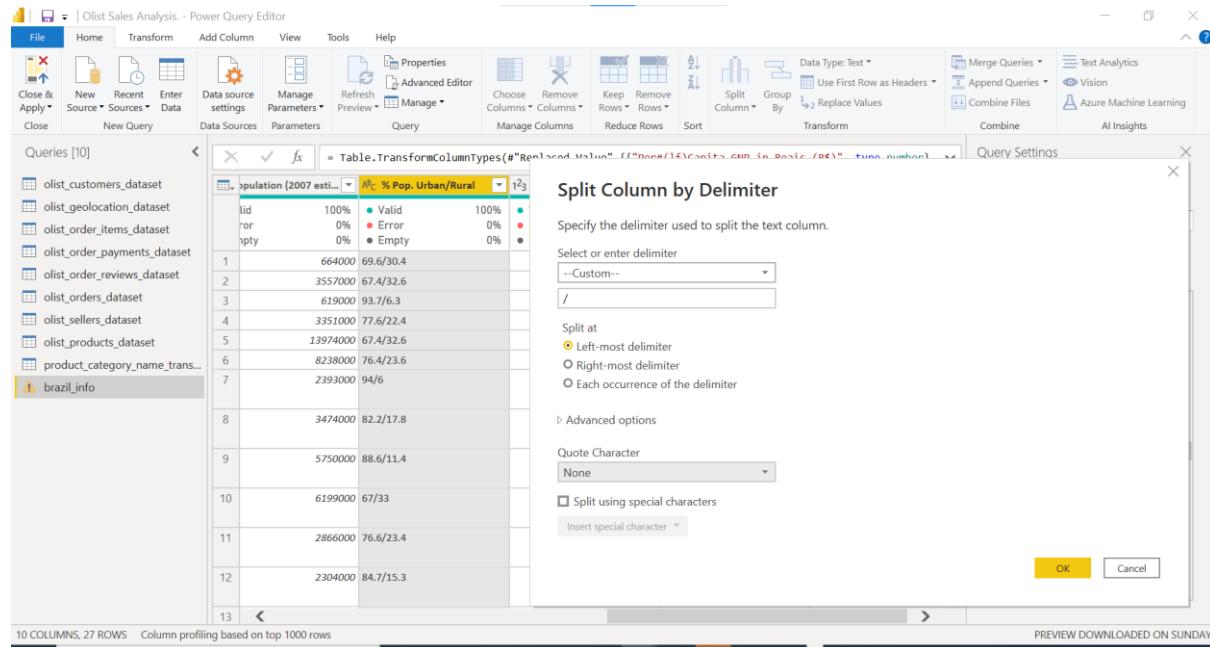
The column '%Pop.Urban/Rural' contains distribution percentage of urban and rural population in the given state. What we need for our model is the percentage of urban population.

So, we use the split function to extract the percentage of urban population in each of the locations listed in the Brazil info dataset by following the steps below.

- From the home tab click the drop-down arrow on the 'split column' and select 'split by delimiter'
- In the window that opens, choose custom and insert '/' which is the character separating the text value we want to split and hit 'ok'.

The screenshot shows the Power Query Editor interface with the 'brazil_info' query selected. In the center, there's a table with several columns. The 'Split Column' button in the ribbon has a dropdown arrow pointing down, revealing a menu. The 'By Delimiter' option is highlighted with a red box and a red arrow pointing to it. The menu also includes other options like 'By Number of Characters', 'By Positions', and 'By Lowercase to Uppercase'. On the right side of the screen, there's a 'Properties' pane showing the query settings and applied steps, with 'Changed Type3' currently selected.

Figure 30: Using 'split column by delimiter' to extract percentage of urban population



2.2.9 Remove Column

The last step create 2 column and we want to delete the unwanted column '%Pop.Urban/Rural.2' and keep '%Pop.Urban/Rural.1' column following the steps below.

- Select the unwanted column
- Click the 'remove column' button in the home tab

Figure 31: Deleting column using the 'remove column' button

The screenshot shows the Power Query Editor interface. In the top ribbon, the 'Transform' tab is selected. On the far right of the ribbon, there is a 'Transform' section containing several icons: 'Choose Columns', 'Remove Columns' (which has a red arrow pointing to it), 'Keep Rows', 'Remove Rows', 'Split Column', 'Group By', 'Data Type: Decimal Number', 'Use First Row as Headers', 'Merge Queries', 'Append Queries', 'Text Analytics', 'Combine Files', 'Azure Machine Learning', and 'AI Insights'. Below the ribbon, the 'Queries [10]' pane lists various datasets. The main preview area shows a table with columns: 'Population (2007 est.)', '1.2 % Pop. Urban/Rural.1', '1.2 % Pop. Urban/Rural.2', 'Number of Municipal...', and 'Per Capita GNP in Re...'. The 'APPLIED STEPS' pane on the right shows a list of steps, with 'Changed Type4' being the last step applied.

2.2.10 Rename Column

We want to rename the '%Pop.Urban/Rural' column created to a more relatable name '

- Right-click on the header of the column and select 'rename'
- Delete existing name and type in desired name and hit enter

Figure 32: Re-naming column with more descriptive name

The screenshot shows the Power Query Editor interface. The 'Transform' tab is selected in the ribbon. The main preview area shows a table with columns: 'Population (2007 est.)', '1.2 % Pop. Urban/Rural', 'Per Capita GNP in Re...', and '1.2 Life Expectancy (200...'. A context menu is open over the second column header ('1.2 % Pop. Urban/Rural'). The menu options include 'Copy', 'Remove', 'Remove Other Columns', 'Duplicate Column', 'Add Column From Examples...', 'Remove Duplicates', 'Remove Errors', 'Change Type', 'Transform', 'Replace Values...', 'Replace Errors...', 'Group By...', 'Fill', 'Unpivot Columns', 'Unpivot Other Columns', 'Unpivot Only Selected Columns', and 'Rename...'. The 'Rename...' option is highlighted with a red box. The 'APPLIED STEPS' pane on the right shows a list of steps, with 'Removed Columns' being the last step applied.

Figure 33: Showing the re-named column 'urban population (%)'

The screenshot shows the Power Query Editor interface with a table containing 10 columns and 27 rows. The columns are labeled: Population (2007 est.), Urban population(%), Number of Municipal..., Per Capita GNP in Re..., and Life Exp. A context menu is open over the 'Urban population(%)' column, with 'Rename' selected. The 'APPLIED STEPS' pane shows the step 'Renamed Columns'.

2.2.11 Rename Tables

To get rid of redundant text in the table name, we remove ‘dataset’ from all the names since we already know we are dealing with datasets. The steps below describe how to rename the first table, and the process was repeated for the other tables.

- Select table title from list of datasets on the left panel
- Right-click and select ‘rename’ as shown in the figure below
- Use backspace to remove ‘_dataset’ from the table title and hit ‘enter’
- Use the ‘close and apply’ tab to load all datasets into Power BI.

The screenshot shows the Power Query Editor interface with the 'olist_customers_dataset' table selected. A context menu is open over the table title, with 'Rename' selected. The 'APPLIED STEPS' pane shows the step 'Filtered Rows'.

Figure 34: Rename the table by deleting '_dataset' from the title

The screenshot shows the Power Query Editor interface with the following details:

- File Bar:** File, Home, Transform, Add Column, View, Tools, Help.
- Queries [10]:** A list of datasets including 'olist_geolocation_dataset', 'olist_order_items_dataset', etc., and 'olist_customers_dataset'.
- Current Query:** 'olist_customers_dataset' (highlighted in yellow).
- Table View:** Shows a preview of the 'customer' table with columns: 'customer_id', 'customer_unique_id', 'customer_zip_code_prefix', and 'customer_city'. The data consists of 999 rows and 5 columns.
- Properties Panel:** Shows the 'Name' field set to 'olist_customers_dataset'.
- Applied Steps Panel:** Shows the 'APPLIED STEPS' section with the 'Changed Type' step applied to the 'customer' column.
- Bottom Status:** '5 COLUMNS, 999+ ROWS' and 'PREVIEW DOWNLOADED AT 18:52'.

Figure 35: 'Close and Apply' to load the renamed datasets into Power BI

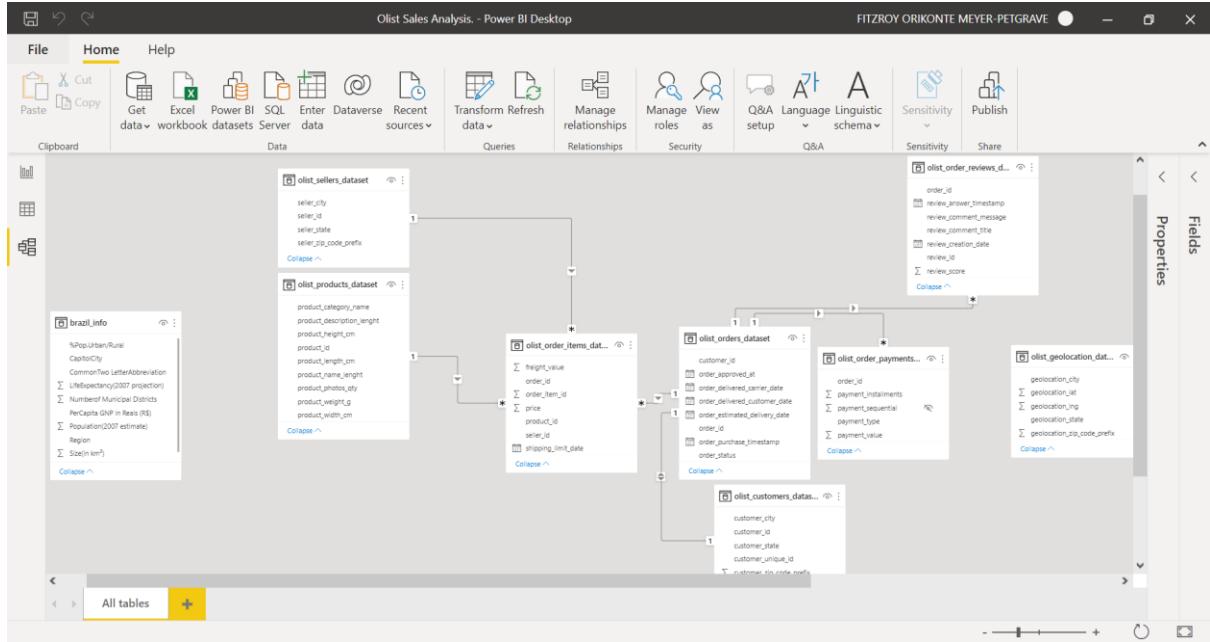
The screenshot shows the Power Query Editor interface with the following details:

- File Bar:** File, Home, Transform, Add Column, View, Tools, Help.
- Queries [10]:** A list of datasets including 'olist_geolocation', 'olist_order_items', etc., and 'olist_customers'.
- Current Query:** 'olist_customers' (highlighted in yellow).
- Table View:** Shows a preview of the 'customer' table with columns: 'customer_id', 'customer_unique_id', 'customer_zip_code_prefix', and 'customer_city'. The data consists of 999 rows and 5 columns.
- Properties Panel:** Shows the 'Name' field set to 'olist_customers'.
- Applied Steps Panel:** Shows the 'APPLIED STEPS' section with the 'Changed Type' step applied to the 'customer' column.
- Bottom Status:** '5 COLUMNS, 999+ ROWS' and 'PREVIEW DOWNLOADED AT 19:17'.

3 DATA MODELLING – STAR SCHEMA

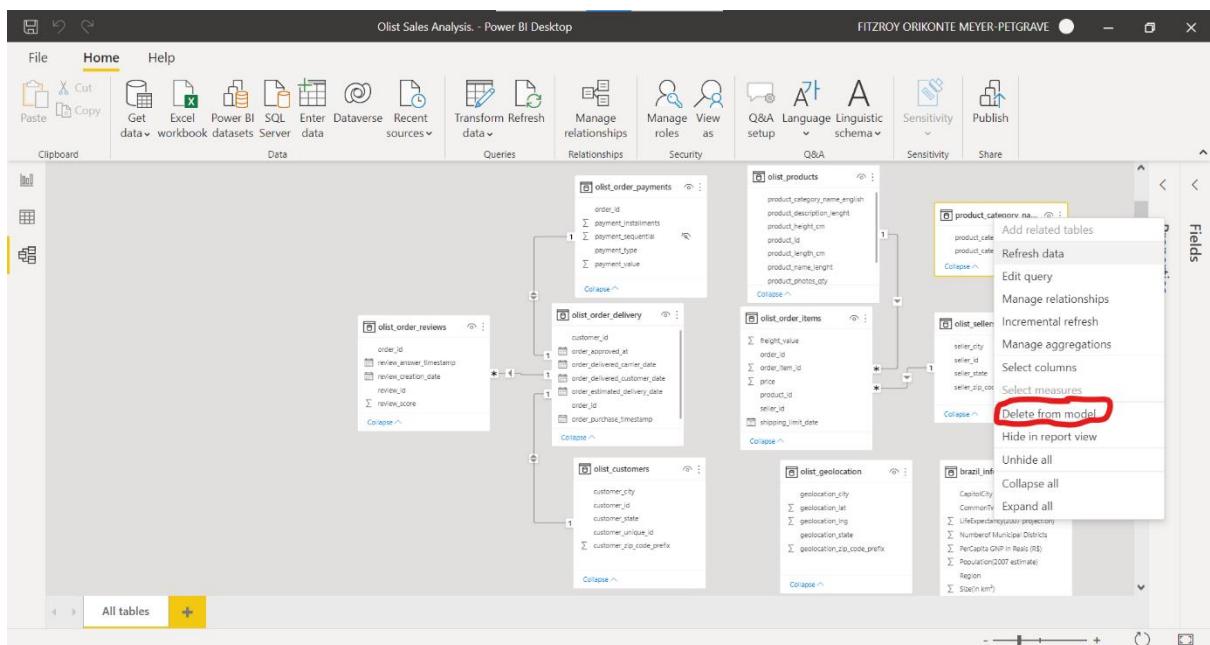
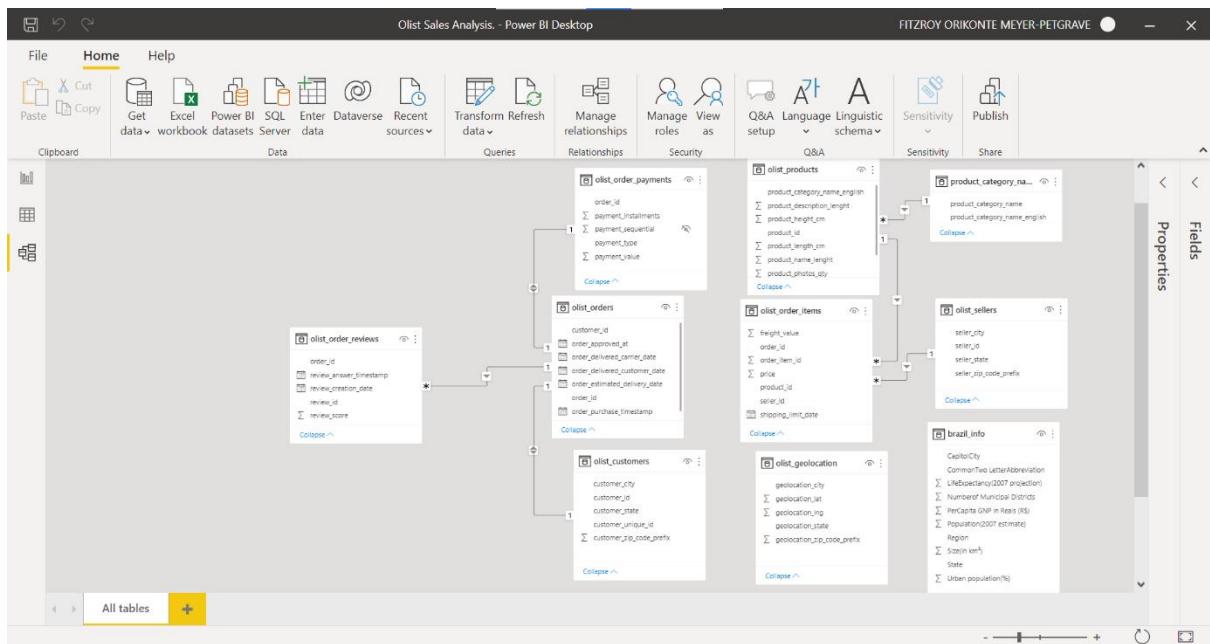
The aim here is to create a model that will have fact tables and dimensions tables as needed. Below is a table showing information about our business intelligence model using star schema method.

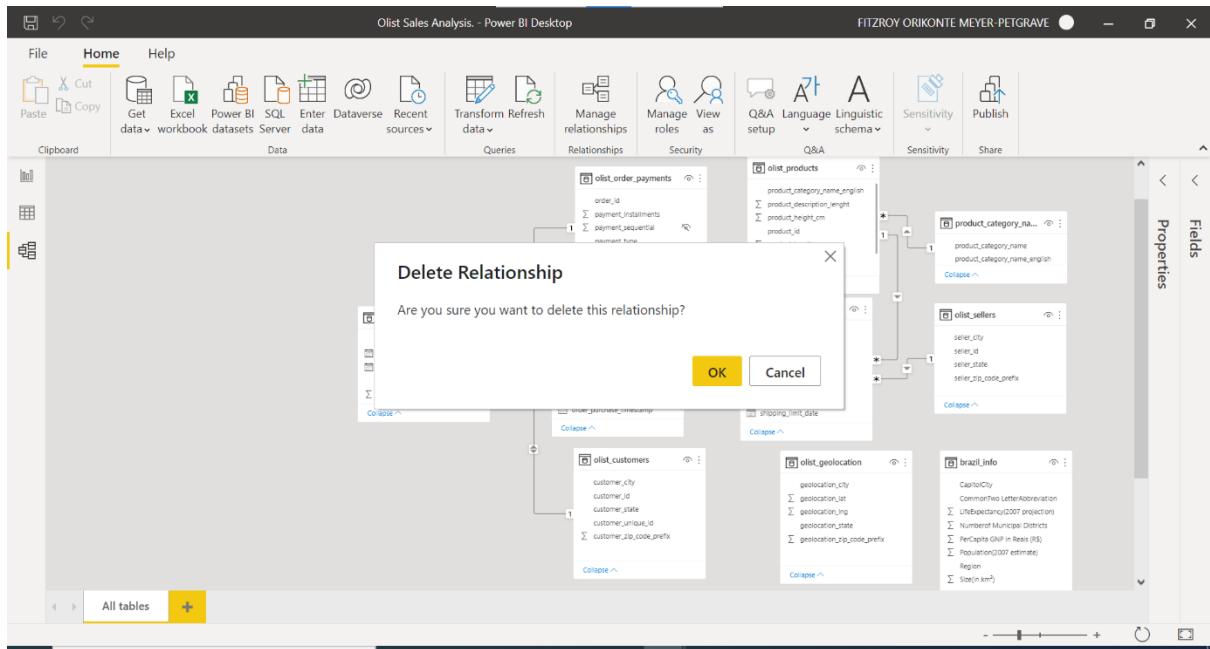
Figure 36: Initial Model showing auto-generated table relationships



3.1 Delete Relationship and re-arrange model

Rearranging the tables to form a star schema with 2 fact tables and 8 dimension tables. Reduce the dimension deleting one table which has already been merged to the product table using the product name. to delete this table, first delete the relationship and rearrange model.





3.2 Create New Relationships

Using 'Manage relationships' to link each of the dimension tables to at least one fact table.

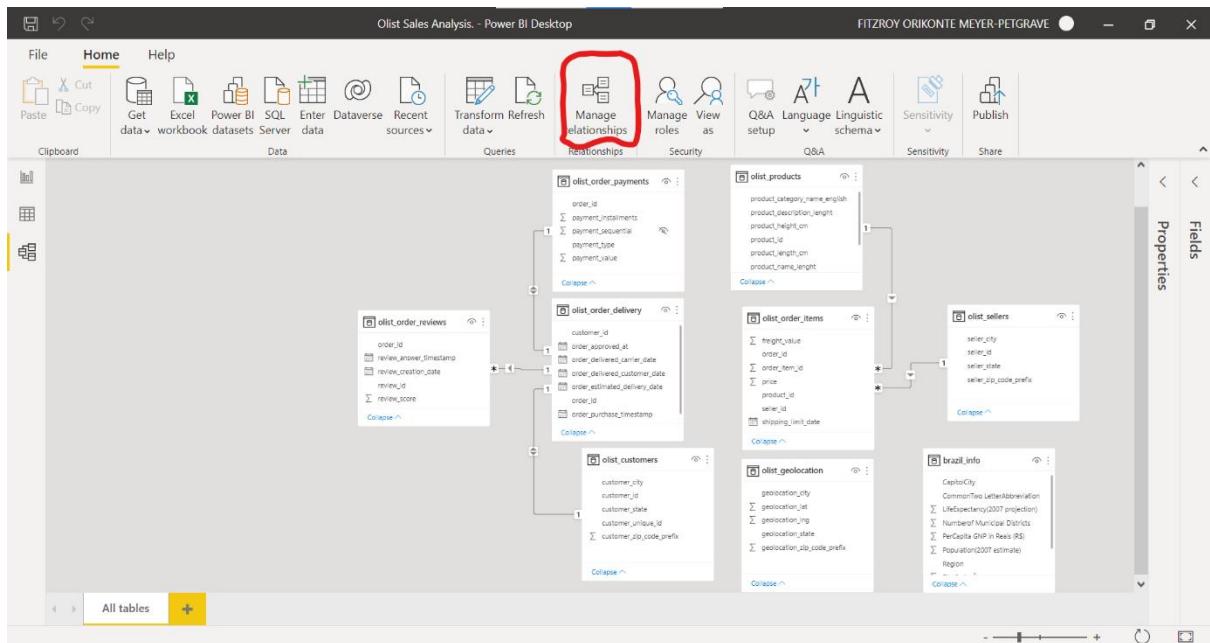


Figure 37: Link the two facts tables by order id

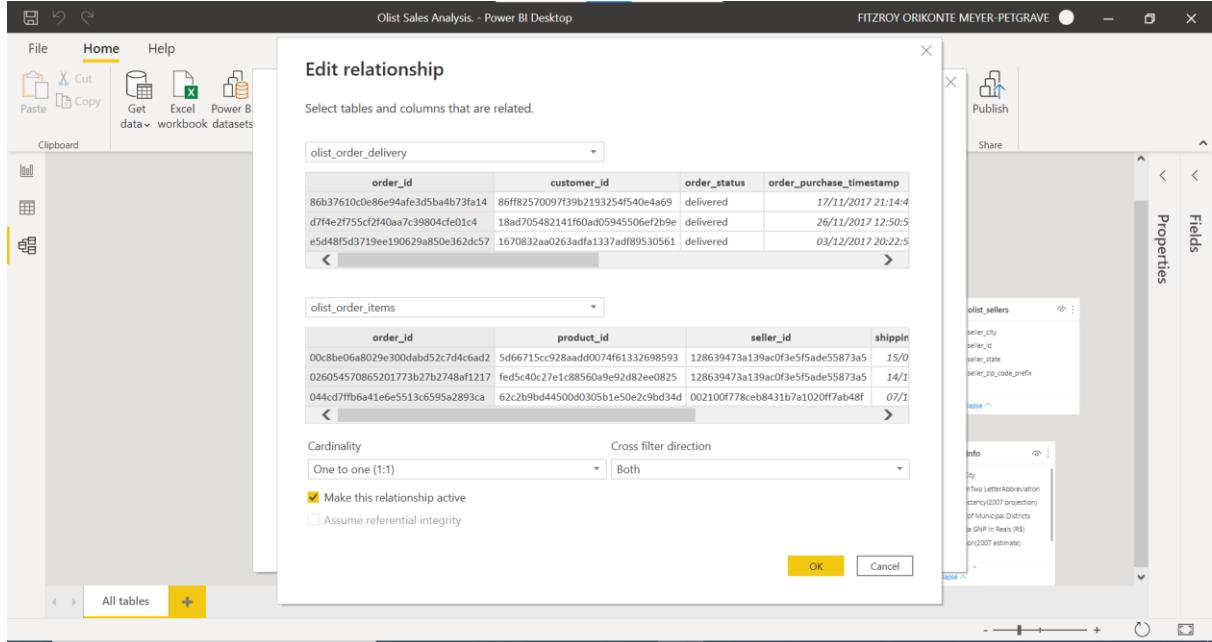


Figure 38: Link customers to geolocation

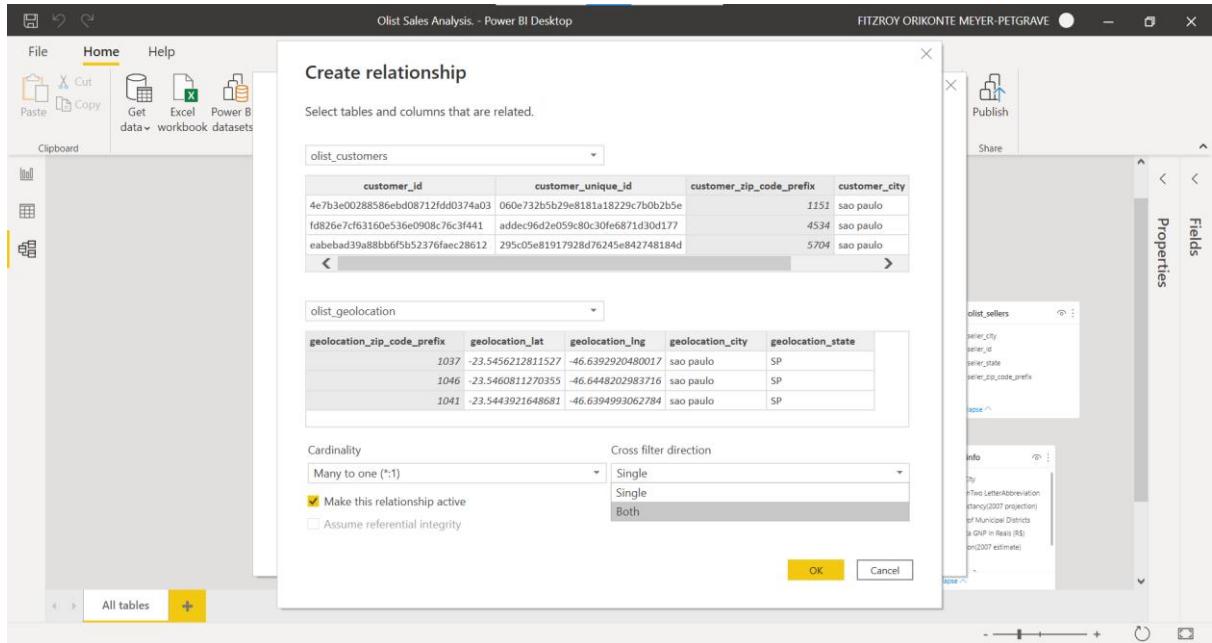


Figure 39: Also link geolocation to Brazil data by state

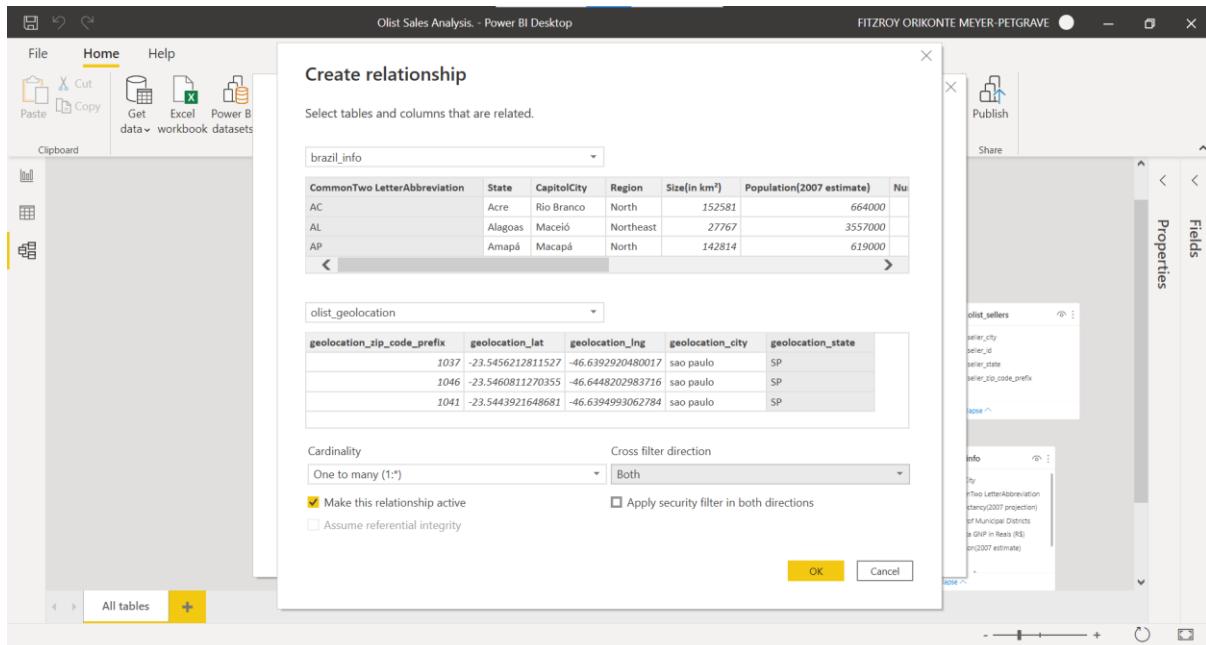
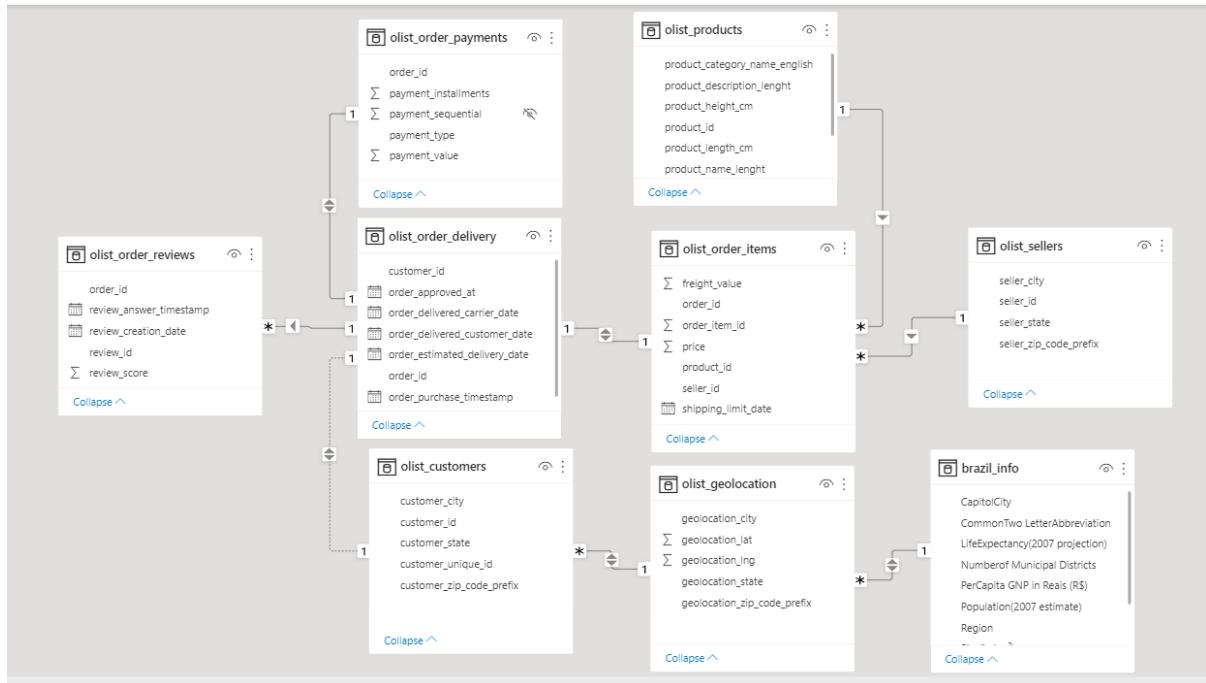
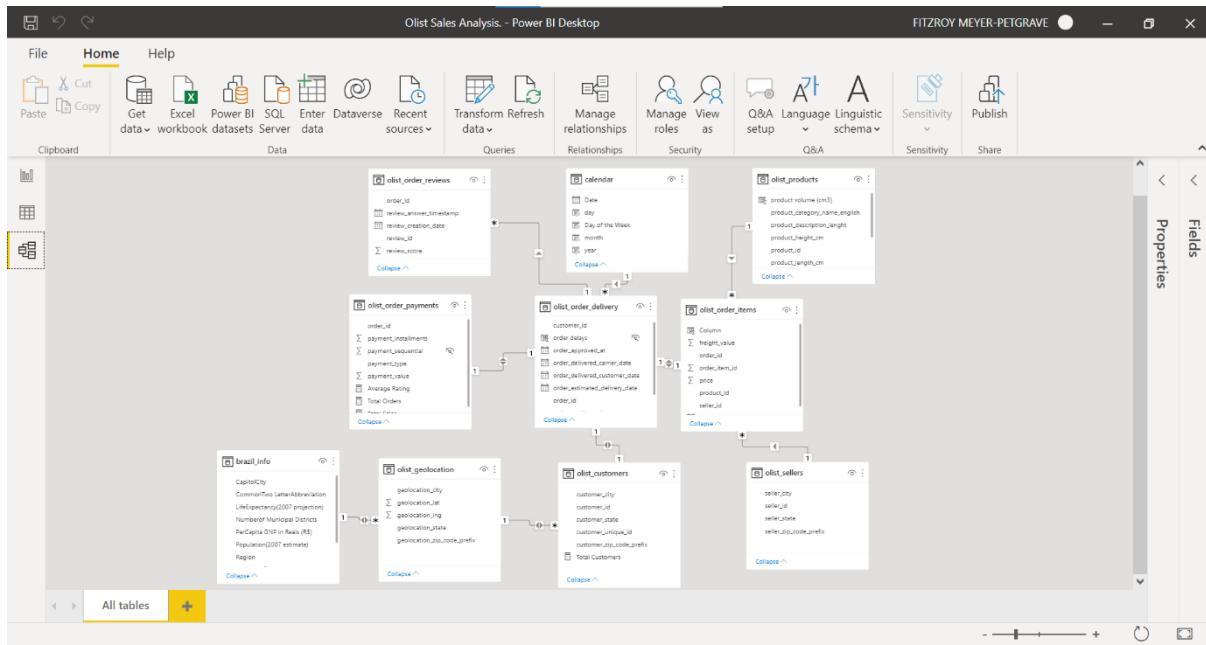


Figure 40: Result of relationships created





3.3 Create Date Table using DAX

Create date table using DAX formular. Maximum is set to the latest date in the review answer timestamp column which is the most recent review taken after the last order.

minimum is also set to the earliest date on order purchase timestamp column in olist order delivery table.

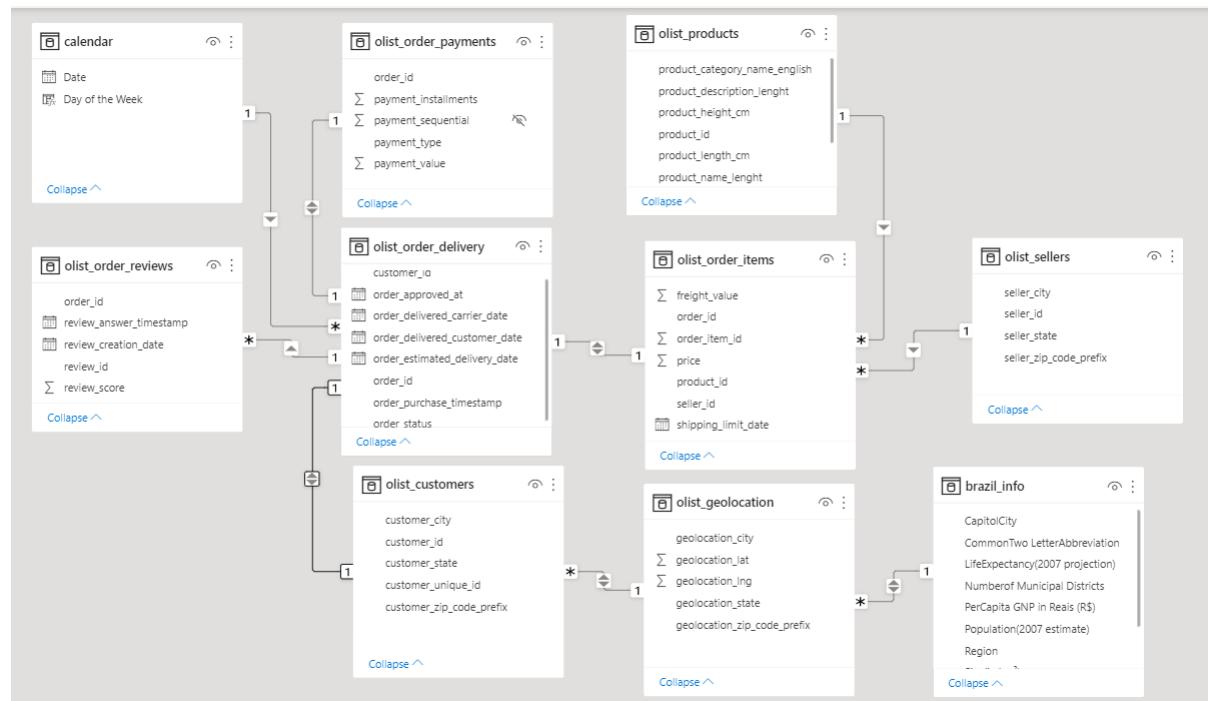
```
1 calendar = CALENDAR(MIN(olist_order_delivery[order_purchase_timestamp]), MAX(olist_order_reviews[review_answer_timestamp]))
```

Figure 41: Dax code for calendar

Figure 42: Create new column for day of the week

```
1 Day of the Week = FORMAT(calendar'[Date], "dddd")
```

The screenshot shows the 'Column tools' ribbon selected in Power BI Desktop. A new column named 'Day of the Week' has been created, with the formula `FORMAT(calendar'[Date], "dddd")`. The data type is set to Text. The column properties pane shows the formula and data type. The Fields pane on the right lists various tables and columns, with 'Day of the Week' highlighted.



MEYER-PETGRAVE, FITZROY- A0384858
TEESSIDE UNIVERSITY

Performance & Analytics Dashboard

BUSINESS INTELLIGENCE REPORT

B. SECTION 2: BUSINESS INTELLIGENCE REPORT

3.4 EXECUTIVE SUMMARY

Basic visualisation of data is no longer enough to meet demand of a highly populated market segment - e commerce. To get ahead of competitors there is need to get more insights about data which will determine how much of that environment you can understand.

This model can allow you to understand your business data from dynamic perspectives, things like how the prosperity of your customer location can affect performance indicators for your business.

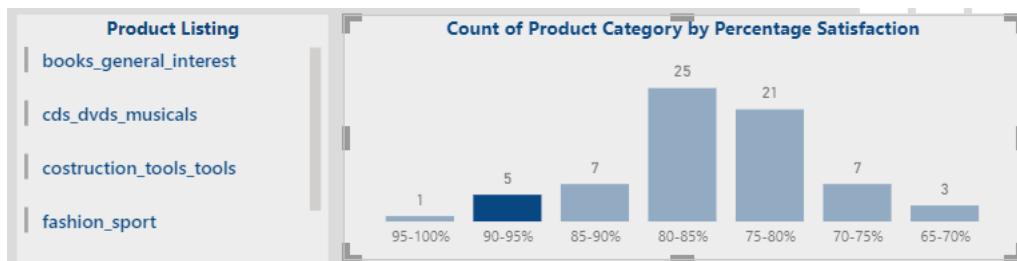
This is a demonstration of a real e commerce business scenario - Olist Brazil Ecommerce platform. It is an original combination of Olist Sales information. The addition of Brazilian city and state information introduced a dynamic perspective to help us understand the customer, the environment and how their preferences affect business growth.

3.4.1 Key Findings

The model is made of 5 sections

1. Overview performance: this gave an overview of the Olist sales performance. This dashboard revealed that business grew from 2016 to 2017 but saw a decline in 2018.
2. Current insights: Insights were derived using calculated measures and columns on the sales and orders information. The comparisons based of products and location revealed the locations and products that generate the most as well as the least revenue. Customer Satisfaction was also analysed to determine which product customers were mostly satisfied with as shown below.

Figure 43: Product category Ranking showing percentage customer satisfaction with category listing



3. New business KPI drivers: Product information like volume and number of photos listed were analysed using AI functions in Power BI to discover the type of influence it has on business growth. Information on Location were also used for this analysis.

3.5 INTRODUCTION

3.5.1 Key Performance Indicators

In the Olist E-commerce platform, revenue is generated from payment for orders made. So, for the business model to be profitable, customers need to increase the number of orders made as well as payments. A clear indicator of business growth is an increase in sales amount, as well as number of orders. However, if customers are unsatisfied, it reduces the chances of having repeat-customers. So, in this case study the key performance indicators would be sales revenue, number of orders, customer satisfaction and number of customers.

Strategic decision-making in business is about evaluating the pros and cons of a situation and developing a stepwise approach to realize your business goals. The business goals in this model would be tailored to its business niche (e-commerce) and its revenue streams based on available data.

The design of this model is made to dynamically explore the available data from Olist as well as that of the Brazilian states where customers reside. It utilises basic charts as well as AI and ML – powered visualisations in Power BI to reveal some patterns in our KPIs over time.

3.5.2 Business Questions Addressed

This model will analyse and reveal useful patterns within the data in ways that are both presentable and actionable. These insights would serve as leads that will help set marketing objectives, design customer loyalty programs and expansion plans.

Using dynamic formulars, functions and special charts, the model was designed to query our data to derive answers to the questions listed below.

1. What are the measures of overall performance metrics?
2. What are the trends in sales and number of orders?
3. How have the KPIs changed over time?
4. What are the products that customers are the most/least satisfied with?
5. Which locations and products generated the most revenue and orders?
6. How are the KPSs affected by – Product Volume; Number of product photos; Population and level of development of customer locations.

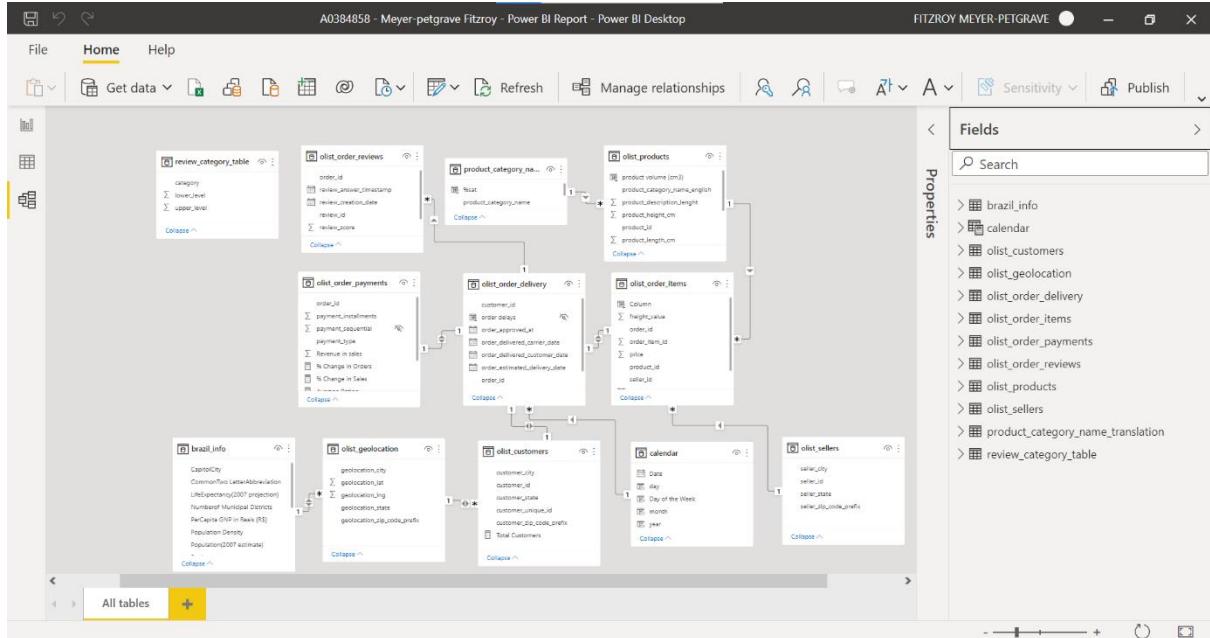
3.5.3 Description of The Data Used in The Model

This dataset is an actual sales data for Olist Ecommerce Company in Brazil which contains information on payments, delivery, reviews, products and customer information, from 2016 to 2018. The company made this dataset available on the company's Kaggle profile which can be accessed via this link from this link: <https://www.kaggle.com/olistbr/brazilian-e-commerce>.

To have a dynamic perspective in this report, a table containing Brazilian state information was introduced into the model directly from this website: https://brazil-help.com/brazilian_states.htm.

The Brazilian state dataset contains description and statistics of all the states including their capital cities, population, and size information.

Figure 44: Model Schema Showing All Tables Used



3.6 FINDING BASED ON ANALYSIS AND EVALUATION (Measures/Tables/Columns)

3.6.1 Performance Analytics

Growth over time (percentage variance) is a measure that indicates an increase or decline in any of the key performance indicators. With a measure of variance, one can tell the growth or decline during the timeframe considered, and by what percentage.

From the data we have, percentage variance calculation can be derived using timestamp with the values of the KPI in consideration (sales, orders, customer satisfaction).

3.6.1.1 Variance Measure

For this report, we have calculated variance using DAX formular in the data window as shown below. These DAX formulars would return the percentage change in sales and orders respectively.

This DAX formular works by using time intelligence to create variables for previous and current values to be compared. The previous values are taken as the last step backward (-1) depending on which date hierarchy is in scope.

The pictures below show the percentage change for sales and orders over time, respectively.

Figure 45: DAX formular for calculating percentage change in Sales Over Time

Figure 46: DAX formular for calculating percentage change in Orders Over Time

3.6.1.2 Year-Over-Year Growth Rate Measure (Sales and Orders)

The Year-over-year (YOY) growth rate measure shows the percentage change during the past twelve months.

This measure is an effective way of looking at growth for two reasons. First, it removes the effects of seasons, and it discerns long-term trends.

This measure is applied to sales and well as orders using the DAX formular shown below.

Figure 47: Sales Revenue (%YoY Growth)

```
1 Revenue in sales Yo% =
2 IF(
3     ISFILTERED('calendar'[Date]),
4     ERROR("Time intelligence quick measures can only be grouped or filtered by the Power BI-provided date hierarchy or primary date column."),
5     VAR __PREV_YEAR =
6         CALCULATE(
7             SUM('olist_order_payments'[Revenue in sales]),
8             DATEADD('calendar'[Date].[Date], -1, YEAR)
9         )
10    RETURN
11    DIVIDE(
12        SUM('olist_order_payments'[Revenue in sales]) - __PREV_YEAR,
13        __PREV_YEAR
14    )
15 )
```

Figure 48: Count of Orders (%YoY Growth)

```
1 count of orders YoY% =
2 IF(
3     ISFILTERED('calendar'[Date]),
4     ERROR("Time intelligence quick measures can only be grouped or filtered by the Power BI-provided date hierarchy or primary date
column."),
5     VAR __PREV_YEAR =
6         CALCULATE(
7             COUNTA('olist_order_items'[order_id]),
8             DATEADD('calendar'[Date].[Date], -1, YEAR)
9         )
10    RETURN
11    DIVIDE(COUNTA('olist_order_items'[order_id]) - __PREV_YEAR, __PREV_YEAR)
12 )
```

3.6.1.3 Month over month percentage growth(measure) – sales and orders

This is like the Year-over-year growth rate measure in that it shows the percentage change over time, however, the Month-Over-Month covers a one-month timeframe.

This measure is applied to sales and well as orders using the DAX formular shown below.

Figure 49: Revenue in Sales (%MoM Growth)

```
1 Revenue in Sales MoM% =
2 IF(
3     ISFILTERED('calendar'[Date]),
4     ERROR("Time intelligence quick measures can only be grouped or filtered by the Power BI-provided date hierarchy or primary date
column."),
5     VAR __PREV_MONTH =
6         CALCULATE(
7             SUM('olist_order_payments'[Revenue in sales]),
8             DATEADD('calendar'[Date].[Date], -1, MONTH)
9         )
10    RETURN
11    DIVIDE(
12        SUM('olist_order_payments'[Revenue in sales]) - __PREV_MONTH,
13        __PREV_MONTH
14    )
15 )
```

Figure 50: Count of Orders (%MoM Growth)

```
1 Count of orders MoM% =
2 IF(
3     ISFILTERED('calendar'[Date]),
4     ERROR("Time intelligence quick measures can only be grouped or filtered by the Power BI-provided date hierarchy or primary date
column."),
5     VAR __PREV_MONTH =
6         CALCULATE(
7             COUNTA('olist_order_items'[order_id]),
8             DATEADD('calendar'[Date].[Date], -1, MONTH)
9         )
10    RETURN
11    DIVIDE(
12        COUNTA('olist_order_items'[order_id]) - __PREV_MONTH,
13        __PREV_MONTH
14    )
15 )
```

3.6.2 Current Insights (Analysing Customer Satisfaction)

One of our KPI measure is customer satisfaction. This is important as it can determine the chance of customer retention. The more satisfied a customer is about product or service, the higher the chance of subsequent purchase.

From the given dataset the metric closely related to customer satisfaction is the customer rating. From rating of 4 and 5 signifies satisfaction, and 1 – 3, dissatisfaction.

Based on the total rating a product category received, we can tell if a customer is satisfied with the product or not.

This will help know which products improve overall customer experience and vice versa.

3.6.2.1 Satisfaction Mapping Based on Ratings (M Language in Power Query)

A new conditional column was created to indicate if a customer was satisfied with a product or not, based on the rating score given during review. Using M language “create conditional column” in Power Query, if the rating 4 and above then map “1” (meaning “satisfied”), and if the rating is less than 4, map 0 (meaning “not satisfied”). The figure below show the computation process.

Figure 51: Creating Conditional Column using M Language in Power Query

The screenshot shows the Microsoft Power Query Editor interface. A modal dialog box titled "Add Conditional Column" is open. Inside the dialog, there is a text input for "New column name" containing "satisfaction". Below this, there is a table with four columns: "Column Name" (containing "review_score"), "Operator" (containing "is greater than or..."), "Value" (containing "A5 123" with a dropdown arrow), and "Output" (containing "ABC 123" with a dropdown arrow). There is also a "Then" field with "1" and an "Else" field with "0". At the bottom right of the dialog are "OK" and "Cancel" buttons. In the background, the main Power Query Editor window shows a list of queries on the left and various tools and options on the top ribbon. A preview pane at the bottom displays a portion of the data with the newly added "satisfaction" column.

3.6.2.2 Ranking Category Table (Using M Language in Power Query)

A new table was created to map out the different categories of satisfaction so that it is easy to highlight/select products based on the average satisfaction customers receive from that product/category. The new table was created in Power Query using M language. The formula specifies a range of values (upper and lower limits) for each category created.

Figure 52: Creating new table with Ranking categories for Customer Satisfaction



```

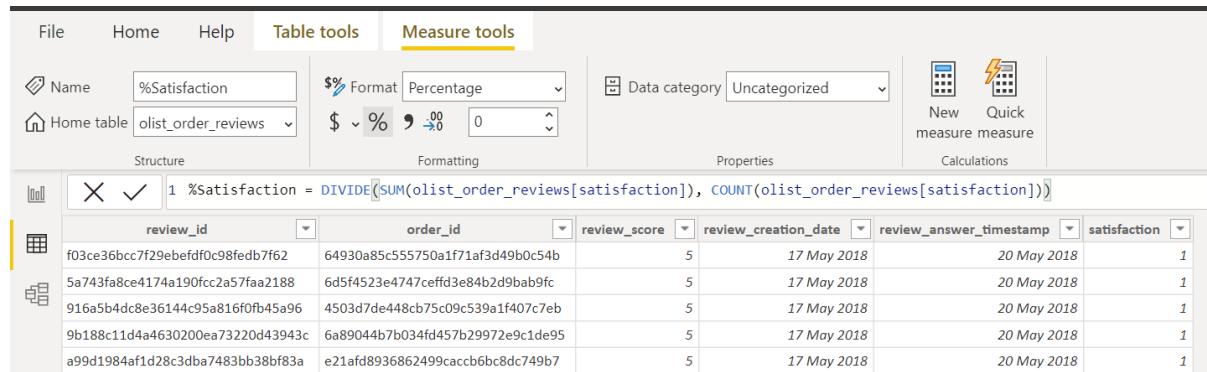
1 let
2     Source = Table.FromRecords({
3         [Category = "65-70%", Lower_Level = 0.65, Upper_Level = 0.70],
4         [Category = "70-75%", Lower_Level = 0.70, Upper_Level = 0.75],
5         [Category = "75-80%", Lower_Level = 0.75, Upper_Level = 0.80],
6         [Category = "80-85%", Lower_Level = 0.80, Upper_Level = 0.85],
7         [Category = "85-90%", Lower_Level = 0.85, Upper_Level = 0.90],
8         [Category = "90-95%", Lower_Level = 0.90, Upper_Level = 0.95],
9         [Category = "95-100%", Lower_Level = 0.95, Upper_Level = 1]
10    }),
11   #"Renamed Columns" = Table.RenameColumns(Source,{{"Category", "category"}, {"Lower_Level", "lower_level"}, {"Upper_Level", "upper_level"}}),
12    #"Changed Type" = Table.TransformColumnTypes(#"Renamed Columns",{{"upper_level", Percentage.Type}, {"lower_level", Percentage.Type}})
13 in
14    #"Changed Type"

```

3.6.2.3 Percentage Satisfaction Measure (Using DAX)

The percentage satisfaction was derived as a measure for product category by using DAX to express the number of “satisfied customers” as a ratio of the total review ratings received. This percentage satisfaction measure is then expressed as a percentage of the total reviews received for each product category. The DAX formula used is shown below.

Figure 53: DAX formula to derive Percentage Satisfaction Measure



The screenshot shows the Power BI DAX editor interface. The top navigation bar has tabs: File, Home, Help, Table tools, and Measure tools. The Measure tools tab is selected. Below the tabs, there are sections for Name (%Satisfaction), Home table (olist_order_reviews), Data category (Uncategorized), and Calculations. The Calculations section contains two buttons: New measure and Quick measure. A preview pane shows the DAX formula:

```

1 %Satisfaction = DIVIDE(SUM(olist_order_reviews[satisfaction]), COUNT(olist_order_reviews[satisfaction]))

```

Below the preview pane is a table view showing review data with columns: review_id, order_id, review_score, review_creation_date, review_answer_timestamp, and satisfaction. The satisfaction column values are all 1.

| review_id | order_id | review_score | review_creation_date | review_answer_timestamp | satisfaction |
|----------------------------------|----------------------------------|--------------|----------------------|-------------------------|--------------|
| f03ce36bcc7f29ebefdf0c98fedb7f62 | 64930a85c555750a1f71af3d49b0c54b | 5 | 17 May 2018 | 20 May 2018 | 1 |
| 5a743fa8ce4174a190fc2a57faa2188 | 6d5f4523e4747cefdf3e84b2d9bab9fc | 5 | 17 May 2018 | 20 May 2018 | 1 |
| 916a5b4dc8e36144c95a816f0fb45a96 | 4503d7de448cb75c09c539a1f407c7eb | 5 | 17 May 2018 | 20 May 2018 | 1 |
| 9b188c11d4a4630200ea73220d43943c | 6a89044b7b034fd457b29972e9c1de95 | 5 | 17 May 2018 | 20 May 2018 | 1 |
| a99d1984af1d28c3dba7483bb38bf83a | e21afd8936862499caccc6bc8dc749b7 | 5 | 17 May 2018 | 20 May 2018 | 1 |

Figure 54: Percentage Satisfaction Measure applied to a New Column

| product_category_name | product_category_name_english | %sat | rank cat |
|--------------------------|-------------------------------|------|----------|
| beleza_saude | health_beauty | 81% | 80-85% |
| informatica_acessorios | computers_accessories | 77% | 75-80% |
| automotivo | auto | 80% | 75-80% |
| cama_mesa_banho | bed_bath_table | 74% | 70-75% |
| moveis_decoracao | furniture_decor | 76% | 75-80% |
| esporte_lazer | sports_leisure | 82% | 80-85% |
| perfumaria | perfumery | 82% | 80-85% |
| utilidades_domesticas | housewares | 80% | 80-85% |
| telefonia | telephony | 76% | 75-80% |
| relogios_presentes | watches_gifts | 78% | 75-80% |
| alimentos_bebidas | food_drink | 85% | 85-90% |
| bebés | baby | 78% | 75-80% |
| papelaria | stationery | 84% | 80-85% |
| tablets_impressao_imagem | tablets_printing_image | 81% | 80-85% |
| brinquedos | toys | 82% | 80-85% |
| telefonia_fixa | fixed_telephony | 75% | 70-75% |
| ferramentas_jardim | garden_tools | 80% | 80-85% |

3.6.2.4 Category Mapping – new column (DAX)

Mapping the percentage satisfaction to predefined ranges (in the category table created) enables one to extract all product that fall within any of those ranges. The DAX formular below was used to create this mapping and specify the given ranges in a new column.

Figure 55: Using DAX to map product categories to Percentage Satisfaction

| product_category_name | product_category_name_english | %sat | rank cat |
|-----------------------------|-------------------------------|------|----------|
| beleza_saude | health_beauty | 81% | 80-85% |
| informatica_acessorios | computers_accessories | 77% | 75-80% |
| automotivo | auto | 80% | 75-80% |
| cama_mesa_banho | bed_bath_table | 74% | 70-75% |
| moveis_decoracao | furniture_decor | 76% | 75-80% |
| esporte_lazer | sports_leisure | 82% | 80-85% |
| perfumaria | perfumery | 82% | 80-85% |
| utilidades_domesticas | housewares | 80% | 80-85% |
| telefonia | telephony | 76% | 75-80% |
| relogios_presentes | watches_gifts | 78% | 75-80% |
| alimentos_bebidas | food_drink | 85% | 85-90% |
| bebés | baby | 78% | 75-80% |
| papelaria | stationery | 84% | 80-85% |
| tablets_impressao_imagem | tablets_printing_image | 81% | 80-85% |
| brinquedos | toys | 81% | 80-85% |
| telefonia_fixa | fixed_telephony | 75% | 70-75% |
| ferramentas_jardim | garden_tools | 80% | 80-85% |
| fashion_bolsas_e_acessorios | fashion_bags_accessories | 82% | 80-85% |

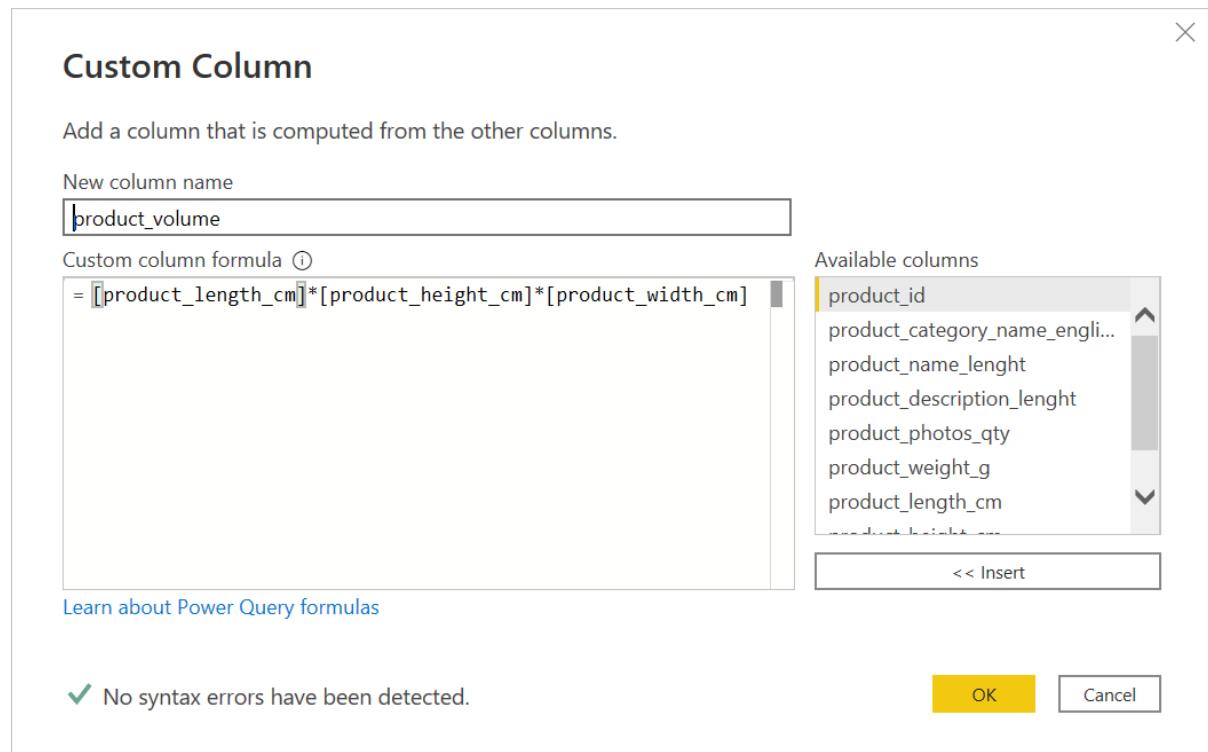
3.6.3 New Business Drivers – Discovering KPI Influencers

To be able to discover new KPI drivers this report makes use of some information about the product and location of customers. However, some metrics need to be derived from the information available in our datasets. The metrics to be derived are the volume of each product and the population density of the state customers are located. This information would be fed into our model to see how they affect sales and orders.

3.6.3.1 Product Volume (Custom column)

Product volume was derived using “custom column calculator” to calculate the volume of a box using basic arithmetic, given that we have length, width and height for each product given in our dataset.

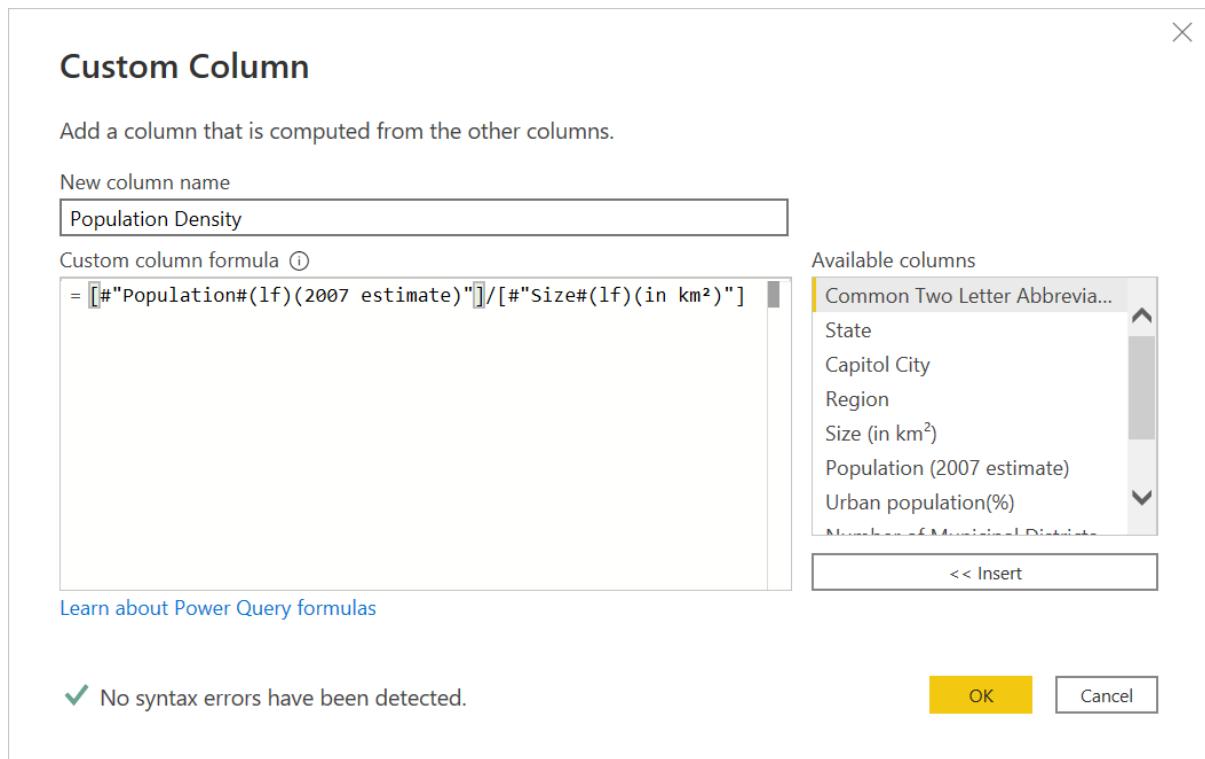
Figure 56: Creating Calculated Custom Columns in Power Query



3.6.3.2 Population density (Custom Column)

Density of a population is basically the number of people living in a unit area of land size. We derive this by dividing the population by the size of the respective state being considered. A DAX formula was used in this report.

Figure 57: Calculating population density in Power Query



3.7 FINDING BASED ON ANALYSIS AND EVALUATION (Charts/Graphs)

POWER BI REPORT – Charts (including one new chart)

3.7.1 Performance Overview

3.7.1.1 What are the measures of overall performance metrics?

From the measures calculated and displayed in the overview dashboard below, over the months from 2016 to 2018, Olist current performance is stated below.

- \$15.74 million dollars in revenue with 3.8% annual growth
- 96,100 customers with average of 77% satisfaction
- 99,000 Orders with 2.27% annual growth

PERFORMANCE & ANALYTICS DASHBOARD

Using Artificial Intelligence in Power BI to analyse business performance and discover new KPI drivers for the Brazilian Ecommerce Platform - OLIST



By

Fitzroy Petgrave

3.7.2 Sales Analytics

The business questions this section addresses are listed below.

- What are the trends in sales?
- How have the Sales changed over time?
- What are the products that customers are the most/least satisfied with?

3.7.2.1 *What are the trends in sales?*

Below is a screen shot showing a column chart of which the bars represent **Sales Revenue** per **day/month/quarter/year** depending on which time frame is in scope. The bars are a good way to represent this because its easy to compare the different time segments and at the same time visualise the actual sales figure for each bar.

Figure 58: Column chart used to represent Total Sales Measure



3.7.2.2 How have the Sales changed over time?

A line and column chart were used to represent the variance measure **percentage change over time**. The green bars show increase while the red show decreasing percentage change over the time in scope.

One key finding is that November and January has the highest increase for the year in scope while September has the highest decrease.

Figure 59: Line and Column Chart Used to Represent Percentage Change Measure

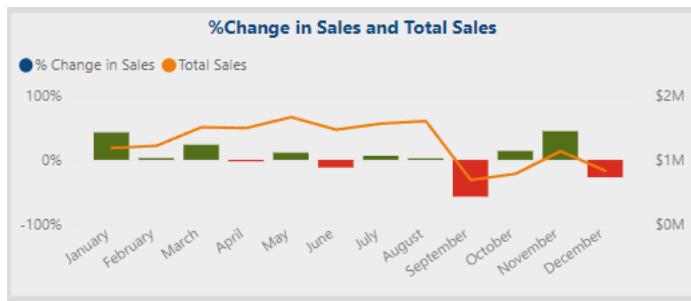


Figure 60: Sales Analytics Dashboard



3.7.3 Orders Analytics

The business questions this section addresses are listed below.

- What are the trends in number of orders?
- How have the Orders changed over time?

3.7.3.1 What are the trends in number of orders?

Below is a screen shot showing a column chart of which the bars represent **Count of Orders** per **day/month/quarter/year** depending on which time frame is in scope. The bars are a good way to represent this because it's easy to compare the different time segments and at the same time visualise the actual sales figure for each bar.

Figure 61: Column chart representing Total Orders



3.7.3.2 How have the Orders changed over time?

A line and column chart were used to represent the variance measure **percentage change over time**. The green bars show increase in orders while the red show decreasing percentage change in orders over the time in scope.

Just like the sales chart, the key finding is that November and January have the highest increase for the year in scope while September has the highest decrease.

Figure 62: %Change in Orders over time



Figure 63: Final Orders Analytics Dashboard

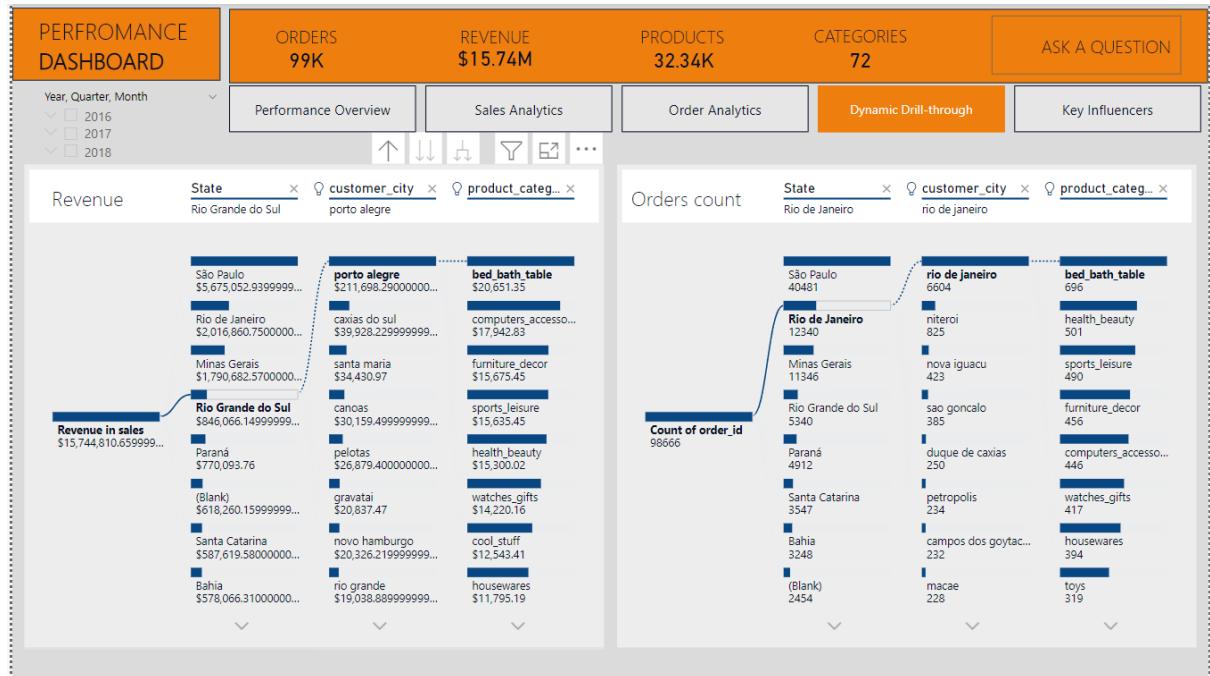


3.7.4 Dynamic Drill-through

3.7.4.1 Which locations and products generated the most revenue and orders?

Power BI decomposition tree is an easy way to break down measures into different components that make it up. In the case study we have used this AI function to split Sale and Revenue into the top contributing states and drilling down into city and product category. The user can interact to break down figure to find out highest and lowest contributors, enabling us to determine which locations and products gave us the most as well as the least orders.

Figure 64: Dynamic drill down dashboard showing Sales and Orders Decomposition Trees



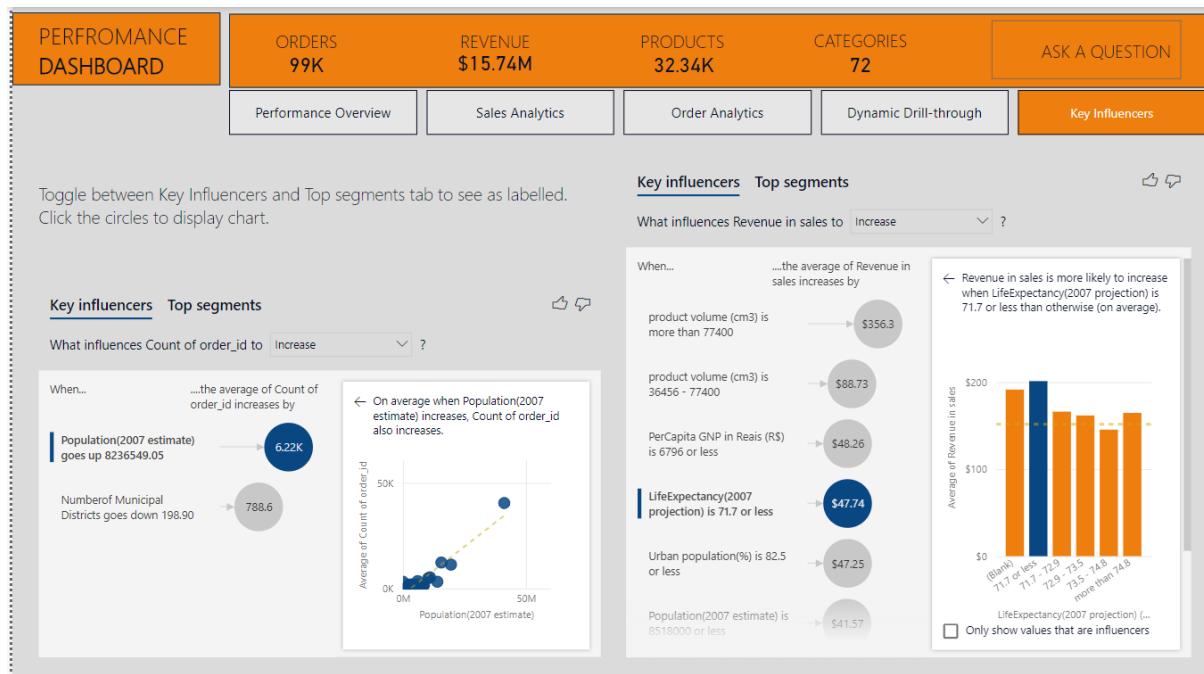
3.7.5 Key Influencers

3.7.5.1 How are the KPSs affected by Product and Location Parameters?

The product parameter being considered is the number of photos for each product and the **Product volume** which has been calculated in power query. While the location parameters are gotten from the Brazilian information table are **Population; GNP per capita, Location size; urban-to-rural ratio, life expectancy and number of municipals.**

The Key Influencers chart in Power BI uses ML algorithms to determine how each of these factors affect sales and number of orders as shown in the chart below.

Figure 65: Key Influencers Chart showing the effect of Product and Location parameters on KPI Measures



3.8 CONCLUSIONS AND RECOMMENDATIONS

3.8.1 Summary of Report

From the analytics it can be concluded that the model was adequate in visualising large data dynamically even with the addition of external datasets.

The top locations that generate the most sales and orders were found to be Sao Paulo, Rio, and Minas. Whereas the top products out of 72 product categories based on number of orders as well as sales revenue are health_beauty, bed_bath_tables, watches_gifts, sports_leisure, housewares and computer accessories.

Some new variables have been discovered to influence sales and orders which are

Product volume, number of product photos, Population, GNP per capita, Location size, urban-to-rural ratio, life expectancy and number of municipals.

3.8.2 Recommendations based on findings

The insights derived with this model can be strategically applied to business processes to drive business growth by applying them to growth activities such as

- Geographical expansion plans

Focus should be on states and cities with low sales and orders, but have the potential in terms of population, which is seen to drive KPI.

- Development of marketing strategies and user retention programs

Customers should be awarded incentives when purchasing products with low satisfaction to increase overall rating and subsequently, customer satisfaction.

- General process monitoring

Sales is inconsistent and hardly predictable. This could be due to the lack of strategic marketing plan. The knowledge of location and products that generate the most/least number in sales and orders should be applied to create a workable system that will generate marketing data for further analysis.

3.8.3 Personal Conclusions

A lot more insights than stated in this report can be derived with this model, especially on the ***Key Influencers'*** dashboard. I will recommend that further studies be carried out to discover parameters that influences ***Customer Satisfaction***, as not enough was covered in this report.

Moreso, I can recommend this model to be used like a pipeline for subsequent/current datasets to generate these insights on demand so that every business decision will come from a lead generated from this Business Intelligence Model.